

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 5, 2006, 19:12:00 ; Search time 2902 seconds
(without alignments)

10479.410 Million cell updates/sec

Title: US-10-664-705-145

Perfect score: 535

Sequence: 1 cggcgcttctgcccag.....cgctgctggtggtcat 535

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_in.*

3: gb_env.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pr.*

9: gb_ro.*

10: gb_sts.*

11: gb_sy.*

12: gb_un.*

13: gb_vi.*

14: gb_hrg.*

15: gb_pi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	535	100.0	928	9	M19262 Rat clathrin
2	471	88.0	982	6	AX401731 Sequence
3	471	88.0	982	6	AX827493 Sequence
4	471	88.0	982	9	M15883 Rat clathrin
5	451.8	84.4	636	8	BT007028 Homo sapi
6	451.8	84.4	636	8	CR536577 Homo sapi
7	451.8	84.4	636	11	AY889907 Synthetic
8	451.8	84.4	636	11	AY892384 Synthetic
9	451.8	84.4	636	11	AY892385 Synthetic
10	451.8	84.4	636	11	BT007759 Synthetic
11	451.8	84.4	1051	8	HUMLCTHB Human lymph
12	451.8	84.4	1099	8	BC006457 Homo sapi
13	451.8	84.4	1164	9	BC006332 Homo sapi
14	442.2	82.7	1009	9	BC070404 Mus muscu
15	429.4	80.3	1033	4	BTLCARL Bos taurus
16	387.8	72.5	1134	8	HUMLCTHB Human brain
17	365.4	68.3	1029	4	BTLCBRB Bos taurus
18	325	60.7	688	6	CQ729113 Sequence

19	323.8	60.5	2615	5	AJ720113
20	261	48.8	738	6	CQ729572 Sequence
21	239.8	44.8	2317	5	BC084464 Xenopus t
22	239.8	44.8	2773	5	CR761042 Xenopus t
23	236.8	44.3	586	9	BC037505 Mus muscu
24	235	43.9	1711	5	BC060412 Xenopus l
25	224	41.9	1751	5	BC077312 Xenopus l
26	223	41.7	1005	4	BTLCARL
27	219.8	41.1	1058	9	BC057660 Mus muscu
28	216.6	40.5	997	5	AJ721067 Gallus ga
29	215	40.2	1034	9	AJ721067 Gallus ga
30	215	40.2	1106	9	M19261 Rat clathri
31	213.4	39.9	657	8	BC087577 Rattus no
32	213.4	39.9	657	11	BT007170 Homo sapi
33	213.4	39.9	657	11	AY890294 Synthetic
34	213.4	39.9	657	11	AY890295 Synthetic
35	213.4	39.9	657	11	AY892771 Synthetic
36	213.4	39.9	657	11	BT008288 Synthetic
37	213.4	39.9	1023	6	CQ724687 Sequence
38	213.4	39.9	1023	8	HUMLCTHA Human lymph
39	213.4	39.9	1055	8	BC019287 Homo sapi
40	197.4	36.9	1103	8	BC009201 Homo sapi
41	187.8	35.1	927	5	BC092127 Xenopus l
42	187.8	35.1	1072	5	CR760072 Xenopus t
43	176.2	32.9	958	6	BC061272 Xenopus t
44	174.2	32.6	400	6	AX305857 Sequence
45	152.2	28.4	1329	5	CQ665260 Sequence
					BC060939 Danio rer

ALIGNMENTS

RESULT 1

RATCBR3	RATCBR3	Rat clathrin light chain (LCB3)	928 bp	mRNA	linear	ROD 27-APR-1993
LOCUS	M19262					
DEFINITION	M19262.1	GI:203360				
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
PUBMED						
COMMENT						
FEATURES						
source						
mRNA						
CDS						
ORIGIN						

```
Query Match      100.0%; Score 535; DB 9; Length 928;
Best Local Similarity 100.0%; Pred. No. 2.4e-111;
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGCGGCTTCTCGGCGCAGCAGAGCGGAGATTGCTGGCATCGAGATGACTCGGGTT 60
Db 99 CGGCGGCTTCTCGGCGCAGCAGAGCGGAGATTGCTGGCATCGAGATGACTCGGGTT 158
Qy 61 TCGGGGACCTTCCGCGCAGCAGGCTGCTCTGCGCAGCCCGGACTCGCGAGCGGGGTG 120
Db 159 TCGGGGACCTTCCGCGCAGCAGGCTGCTCTGCGCAGCCCGGACTCGCGAGCGGGGTG 218
Qy 121 GTTCGGAGGACATGGGAGCTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGGCTG 180
Db 219 GTTCGGAGGACATGGGAGCTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGGCTG 278
Qy 181 CCGATGGCTACGCTGCGGCTGCGGAGGCTGCGGAGGCTGCGGAGGCTTAACGGGCTG 240
Db 279 CCGATGGCTACGCTGCGGCTGCGGAGGCTGCGGAGGCTGCGGAGGCTTAACGGGCTG 338
Qy 241 GCAAGTGGAGAGGAGCAGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 300
Db 339 GCAAGTGGAGAGGAGCAGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 398
Qy 301 CCGAACAGGAGTGGCGGAGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 360
Db 399 CCGAACAGGAGTGGCGGAGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 458
Qy 361 GTGAACAGGTTGAGAGGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 420
Db 459 GTGAACAGGTTGAGAGGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 518
Qy 421 AATCAAGGAGGAGCAGGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 480
Db 519 AATCAAGGAGGAGCAGGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 578
Qy 481 ACCCTAAGAGCAGCAAGCAATGTAAGACGTGTCGCGGCTCGGCTCGGCTCAT 535
Db 579 ACCCTAAGAGCAGCAAGCAATGTAAGACGTGTCGCGGCTCGGCTCGGCTCAT 633

RESULT 2
AX401731
LOCUS      AX401731      982 bp      DNA      linear      PAT 06-JUN-2002
DEFINITION Sequence 1407 from Patent WO0210453.
ACCESSION AX401731
VERSION    AX401731.1 GI:21337911
KEYWORDS   Rattus norvegicus (Norway rat)
SOURCE     Rattus norvegicus
ORGANISM   Rattus norvegicus
REFERENCE 1
AUTHORS    Mendrick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and
            Elashoff,M.R.
TITLE      Molecular toxicology modeling
JOURNAL    Patent: WO 0210453-A 1407 07-FEB-2002;
            Gene Logic, Inc. (US)
FEATURES   Location/Qualifiers
            source
              1..982
              /organism="Rattus norvegicus"
              /mol_type="unassigned DNA"
              /db_xref="taxon:10116"
              /note="EMBL/GenBank Accession No. M15883"
ORIGIN
Query Match      88.0%; Score 471; DB 6; Length 982;
Best Local Similarity 90.8%; Pred. No. 8.8e-97;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

Qy 1 CGGCGGCTTCTCGGCGCAGCAGGAGCGGAGATTGCTGGCATCGAGATGACTCGGGTT 60
Db 99 CGGCGGCTTCTCGGCGCAGCAGGAGCGGAGATTGCTGGCATCGAGATGACTCGGGTT 158
Qy 61 TCGGGGACCTTCCGCGCAGCAGGCTGCTCTGCGCAGCCCGGACTCGCGAGCGGGGTG 120
Db 159 TCGGGGACCTTCCGCGCAGCAGGCTGCTCTGCGCAGCCCGGACTCGCGAGCGGGGTG 218
Qy 121 GTTCGGAGGACATGGGAGCTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGGCTG 180
Db 219 GTTCGGAGGACATGGGAGCTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGGCTG 278
Qy 181 CCGATGGCTACGCTGCGGCTGCGGAGGCTGCGGAGGCTGCGGAGGCTTAACGGGCTG 240
Db 279 CCGATGGCTACGCTGCGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 300
Db 339 GCAAGTGGAGAGGAGCAGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 398
Qy 301 CCGAACAGGAGTGGCGGAGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 360
Db 399 CCGAACAGGAGTGGCGGAGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 458
Qy 361 GTGAACAGGTTGAGAGGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 420
Db 459 GTGAACAGGTTGAGAGGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 518
Qy 421 AATCAAGGAGGAGCAGGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 480
Db 519 AATCAAGGAGGAGCAGGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 578
Qy 481 ACCCTAAGAGCAGCAAGCAATGTAAGACGTGTCGCGGCTCGGCTCGGCTCAT 535
Db 579 ACCCTAAGAGCAGCAAGCAATGTAAGACGTGTCGCGGCTCGGCTCGGCTCAT 633

RESULT 3
AX827493
LOCUS      AX827493      982 bp      DNA      linear      PAT 12-DEC-2003
DEFINITION Sequence 227 from Patent EP1344834.
ACCESSION AX827493
VERSION    AX827493.1 GI:39837681
KEYWORDS   Rattus norvegicus (Norway rat)
SOURCE     Rattus norvegicus
ORGANISM   Rattus norvegicus
REFERENCE 1
AUTHORS    Boess,F., Suter-Dick,L. and Wolf,D.
TITLE      Methods for the toxicity prediction of a compound
JOURNAL    Patent: EP 1344834-A 227 17-SEP-2003;
            F. HOPFMAN-IA ROCHE AG (CH)
FEATURES   Location/Qualifiers
            source
              1..982
              /organism="Rattus norvegicus"
              /mol_type="unassigned DNA"
              /db_xref="taxon:10116"
ORIGIN
Query Match      88.0%; Score 471; DB 6; Length 982;
Best Local Similarity 90.8%; Pred. No. 8.8e-97;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

Qy 1 CGGCGGCTTCTCGGCGCAGCAGGAGCGGAGATTGCTGGCATCGAGATGACTCGGGTT 60
Db 99 CGGCGGCTTCTCGGCGCAGCAGGAGCGGAGATTGCTGGCATCGAGATGACTCGGGTT 158
Qy 61 TCGGGGACCTTCCGCGCAGCAGGCTGCTCTGCGCAGCCCGGACTCGCGAGCGGGGTG 120
Db 159 TCGGGGACCTTCCGCGCAGCAGGCTGCTCTGCGCAGCCCGGACTCGCGAGCGGGGTG 218
Qy 121 GTTCGGAGGACATGGGAGCTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGGCTG 180
Db 219 GTTCGGAGGACATGGGAGCTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGGCTG 278
Qy 181 CCGATGGCTACGCTGCGGCTGCGGAGGCTGCGGAGGCTGCGGAGGCTTAACGGGCTG 240
Db 279 CCGATGGCTACGCTGCGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 300
Db 339 GCAAGTGGAGAGGAGCAGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 398
Qy 301 CCGAACAGGAGTGGCGGAGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 360
Db 399 CCGAACAGGAGTGGCGGAGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 458
Qy 361 GTGAACAGGTTGAGAGGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 394
Db 459 GTGAACAGGTTGAGAGGAGGCTGCGGAGGCTGCGGAGGCTTGGATGCTCGAAGGTGA 518
Qy 395 -----GGCATCGGAGAGGCTTTTGTGAAAGATCCA 426
Db 519 AGCCAGATGCTGATACCATTTGGCTATGTGGCATCGGAGAGGCTTTTGTGAAAGATCCA 578
Qy 427 AGGAGGAGACCCAGGACACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAACCCCTA 486
Db 579 AGGAGGAGACCCAGGACACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAACCCCTA 638
Qy 487 AGAGCAGCAAGCAATGTAAGACGTGTCGCGGCTCGGCTCGGCTCAT 535
Db 639 AGAGCAGCAAGCAATGTAAGACGTGTCGCGGCTCGGCTCGGCTCAT 687

Query Match      88.0%; Score 471; DB 6; Length 982;
Best Local Similarity 90.8%; Pred. No. 8.8e-97;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;
```

```
Db 159 TCGGGGACCTGCGCCAGCCAGGTGGGCTCTGCGCAGCCCGGACTCGGAGCGGGGTG 218
Qy 121 GTTCGGAGCATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACCGGGCCTG 180
Db 219 GTTCGGAGCATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACCGGGCCTG 278
Qy 181 CCGATGGCTACGCTCGGATTCGCCAGCGGACAGGTTCGACTCAGGAGCCTCAGAGCATCC 240
Db 279 CCGATGGCTACGCTCGGATTCGCCAGCGGACAGGTTCGACTCAGGAGCCTCAGAGCATCC 338
Qy 241 GCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 339 GCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 398
Qy 301 CCGAACAGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 399 CCGAACAGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
Qy 361 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 394
Db 459 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 518
Qy 395 -----GGCATCGGAAGAGAGGCTTTTGTGAAGAGATCCA 426
Db 519 AGCCAGATGCTGATACCATTTGCTATGTGGCATCGGAGAGGCTTTTGTGAAGAGATCCA 578
Qy 427 AGGAGGAGACCCAGGACACAGAGTGGGAGAGAGTGGCCAGAGTGTGTGACTTCAACCTTA 486
Db 579 AGGAGGAGACCCAGGACACAGAGTGGGAGAGAGTGGCCAGAGTGTGTGACTTCAACCTTA 638
Qy 487 AGAGCAGCAGCAATGTAAAGAGTGTCCCGCTCGGCTCGGTGCTCAT 535
Db 639 AGAGCAGCAGCAATGTAAAGAGTGTCCCGCTCGGCTCGGTGCTCAT 687
```

```
RESULT 4
RATCBR2 Rat clathrin light chain (LCB2) mRNA linear ROD 27-APR-1993
LOCUS M15883
DEFINITION alternative splicing; clathrin.
ACCESSION Rattus norvegicus (Norway rat)
VERSION Rattus norvegicus
KEYWORDS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
SOURCE Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
ORGANISM Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 982)
AUTHORS Kirchhausen, T., Scarnato, P., Harrison, S.C., Monroe, J.J., Chow, E.P.,
Mattaliano, R.J., Ramachandran, K.L., Smart, J.E., Ahn, A.H. and
Brosius, J.
TITLE Clathrin light chains LCA and LCB are similar, polymorphic, and
share repeated heptad motifs
JOURNAL Science 236 (4799), 320-324 (1987)
PUBMED 3563513
COMMENT Original source text: Rat brain, cDNA to mRNA, clones pPB5-2,
pV12ab, lambda-ACS2-1, and lambda-EC8-4.
Draft entry and computer-readable sequence for [1] kindly provided
by T. Kirchhausen, 05-OCT-1987.
```

```
FEATURES
source Location/Qualifiers
1..982
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
<1..982
/product="LCB2 mRNA"
29..718
/notes="clathryn light chain (LCB2)"
/codon_start=1
/protein_id="AAA40890.1"
/db_xref="GI:203359"
/translation="MAEDFGFSSSEGAPEAAEDPAAFLAQSEIAGTNDGSGF
```

```
ORIGIN 74 bp upstream of AVAL site.
Query Match 88.0%; Score 471; DB 9; Length 982;
Best Local Similarity 90.8%; Pred. No. 8.8e-97;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;
```

```
Qy 1 CGGCGGCTTCTCGCCAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Db 99 CGGCGGCTTCTCGCCAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 158
Qy 61 TCGGGGACCTGCGCCAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 159 TCGGGGACCTGCGCCAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 218
Qy 121 GTTCGGAGCATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACCGGGCCTG 180
Db 219 GTTCGGAGCATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACCGGGCCTG 278
Qy 181 CCGATGGCTACGCTCGGATTCGCCAGCGGACAGGTTCGACTCAGGAGCCTCAGAGCATCC 240
Db 279 CCGATGGCTACGCTCGGATTCGCCAGCGGACAGGTTCGACTCAGGAGCCTCAGAGCATCC 338
Qy 241 GCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 339 GCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 398
Qy 301 CCGAACAGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 399 CCGAACAGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
Qy 361 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 394
Db 459 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 518
Qy 395 -----GGCATCGGAAGAGAGGCTTTTGTGAAGAGATCCA 426
Db 519 AGCCAGATGCTGATACCATTTGCTATGTGGCATCGGAGAGGCTTTTGTGAAGAGATCCA 578
Qy 427 AGGAGGAGACCCAGGACACAGAGTGGGAGAGAGTGGCCAGAGTGTGTGACTTCAACCTTA 486
Db 579 AGGAGGAGACCCAGGACACAGAGTGGGAGAGAGTGGCCAGAGTGTGTGACTTCAACCTTA 638
Qy 487 AGAGCAGCAGCAATGTAAAGAGTGTCCCGCTCGGCTCGGTGCTCAT 535
Db 639 AGAGCAGCAGCAATGTAAAGAGTGTCCCGCTCGGCTCGGTGCTCAT 687
```

```
RESULT 5
BT007028 Homo sapiens clathrin, light polypeptide (Lcb) mRNA linear PRI 13-MAY-2003
LOCUS BT007028
DEFINITION Homo sapiens clathrin, light polypeptide (Lcb) mRNA, complete cds.
ACCESSION BT007028
VERSION BT007028.1 GI:30582894
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 636)
AUTHORS Kalnine, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S.,
Koundinya, M., Raphael, J., Moreira, D., Kelley, T., LaBaer, J., Lin, Y.,
Phelan, M. and Farmer, A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 636)
AUTHORS Kalnine, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S.,
Koundinya, M., Raphael, J., Moreira, D., Kelley, T., LaBaer, J., Lin, Y.,
Phelan, M., and Farmer, A.
```



```

Qy 481 ACCCTAAGACGACGAAGCAATGTAAAGACGTGTCCCGCTCGCGTCTCGTCTCAT 535
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 ACCCAAGACGACGAAGCAAGTGTGTCCCGCTCGCGTCTCGTCTCAT 605

RESULT 8
LOCUS AY892384 636 bp mRNA linear SYN 29-MAR-2005
DEFINITION Synthetic construct Homo sapiens clone FLH025388.01L clathrin light
polypeptide (CLTB) mRNA, partial cds.
ACCESSION AY892384
VERSION AY892384.1 GI:60653212
KEYWORDS Human ORF project.
SOURCE synthetic construct
ORGANISM
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 636)
Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
Cloning of human full-length CDS in Creator (TM) recombinational
vector system
2 (bases 1 to 636)
Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
Direct Submission
Submitted (05-JAN-2005) Biological Chemistry and Molecular
Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
Cambridge, MA 02141, USA
This CDS clone is a part of a collection of human full-length
expression clones generated by Harvard Institute of Proteomics.
This ORF clone has been cloned without stop-codon (to allow fusion
with C-terminal tag). The CDS has been directionally cloned using
BD In-Fusion(TM) cloning system between the SalI and HindIII sites
of the pDNR-Dual vector. Additional sequences in the clone: 'ACC'
after SalI site and before 'ATG' to provide Kozak consensus
sequence; 'GG' after last codon and before HindIII site to maintain
reading frame. Each clone is clonally isolated and full-length
sequence-verified.
FEATURES
source
Location/Qualifiers
1..636
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clones="FLH025388.01L"
/lab_host="Escherichia coli DH5alpha T1 resistant"
/note="derived from MGC template"
1..>636
/gene="CLTB"
1..>636
/gene="CLTB"
/note="Lcb"
/codon_start=1
/tranl_table=11
/product="clathrin light polypeptide"
/protein_id="AA29301.1"
/db_xref="GI:60653213"
/translation="MADDFGFFSSSGAPEAABEDPAAFLAQOSEIAGIENDEGF
GAPGSHAAPAGPPTSGAGSDMGTTVNGDVGQANGPADGVAALAAQADRLTQEPES
IRKWRERQRLQELDAASKVTEQREKAKDLLEWNQKQSOVKNNKINRASEA
FVKSEKETPGTEWKEVAQLCDENFPKSSKOCKDVSRILRSVLMSLKPTLSRL"

ORIGIN
Query Match 84.4%; Score 451.8; DB 11; Length 636;
Best Local Similarity 90.3%; Pred. No. 2e-92;
Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 1 CGGCCGCTTCTCGCCCGACGAGCGGAGATTGCTGGCATCGAAGTACCTCGGTT 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 CGGCCGCTTCTCGCCCGACGAGCGGAGATTGCTGGCATCGAAGTACCTCGGTT 130

```

```

Qy 61 TCGGGGCACCTGCGCCGACGACGAGTGGCTCTGCGCAGCCGCGACTCGCGAGCGGGGTG 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
131 TCGGGGCACCTGCGCCGACGACGATGCGGCCCGCGCAGCCGCGGCCCCCAGGATGGGGCTG 190
Qy 121 GTTCGAGGACATGGGGGACTACAGTCAATGAGAGATGTGTTTCAGGAGGCTTAACGGGGCTG 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
191 GTTCTGAGGACATGGGGGACCCACAGTCAATGAGAGATGTGTTTCAGGAGGCCAACGGTCTG 250
Qy 181 CCGATGGCTACGCTGCGATTTGCCAGCGGACGAGGTGTGACTCAGGAGCCTGAGAGCATCC 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 CTGATGGCTACGCGAGCCATTGCCAGGCTGACACGGCTGACCCAGGAGCCTGAGAGCATCC 310
Qy 241 GCAAGTGGAGAGGAGGACAGAGAAGAGGCTGCAGGAGTTGGATGCTGCTCGAAGGTGA 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
311 GCAAGTGGAGAGGAGGACAGAGAAACGGCTGCAGAGCTGGATGCTGCATCTAAGGTCA 370
Qy 301 CCGAACAGGAGTGGCGGGAGAGGCCAAAGAACCTTGGAGGAGTGGAAACAGCGCCAAA 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 CGGAACAGGAATGGCGGGAGAGGCCAAAGAACCTTGGAGGAGTGGAAACAGCGCCAGA 430
Qy 361 GTGAACAGGTTGAGAGAACAGATCAACAAACAGGGCATCGAAGAGGCTTTTGTGAAG 420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
431 GTGAACAAAGTAGAGAGAACAGATCAACAAACCGGGCATCCGAGGAGGCTTTTCGTGAAG 490
Qy 421 AATCCAAGGAGGAGACCCAGGACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCA 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
491 AATCCAAGGAGGAGACCCAGGACAGAGTGGGAGAGGTGGCCAGCTATGTGACTTCA 550
Qy 481 ACCCTAAGACGACGAAGCAATGTAAAGACGTGTCCCGCTCGCGTCTCGTCTCAT 535
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 ACCCAAGACGACGAAGCAAGTGTGTCCCGCTCGCGTCTCGTCTCAT 605

RESULT 9
AY892385
LOCUS AY892385 636 bp mRNA linear SYN 29-MAR-2005
DEFINITION Synthetic construct Homo sapiens clone FLH025389.01L clathrin light
polypeptide (CLTB) mRNA, partial cds.
ACCESSION AY892385
VERSION AY892385.1 GI:60653214
KEYWORDS Human ORF project.
SOURCE synthetic construct
ORGANISM
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 636)
Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
Cloning of human full-length CDS in Creator (TM) recombinational
vector system
Unpublished
2 (bases 1 to 636)
Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
Direct Submission
Submitted (05-JAN-2005) Biological Chemistry and Molecular
Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
Cambridge, MA 02141, USA
This CDS clone is a part of a collection of human full-length
expression clones generated by Harvard Institute of Proteomics.
This ORF clone has been cloned without stop-codon (to allow fusion
with C-terminal tag). The CDS has been directionally cloned using
BD In-Fusion(TM) cloning system between the SalI and HindIII sites
of the pDNR-Dual vector. Additional sequences in the clone: 'ACC'
after SalI site and before 'ATG' to provide Kozak consensus
sequence; 'GG' after last codon and before HindIII site to maintain
reading frame. Each clone is clonally isolated and full-length
sequence-verified.
FEATURES
Location/Qualifiers
1..636
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clones="FLH025389.01L"
/lab_host="Escherichia coli DH5alpha T1 resistant"
/note="derived from MGC template"
1..>636
/gene="CLTB"
1..>636
/gene="CLTB"
/note="Lcb"
/codon_start=1
/tranl_table=11
/product="clathrin light polypeptide"
/protein_id="AA29301.1"
/db_xref="GI:60653213"
/translation="MADDFGFFSSSGAPEAABEDPAAFLAQOSEIAGIENDEGF
GAPGSHAAPAGPPTSGAGSDMGTTVNGDVGQANGPADGVAALAAQADRLTQEPES
IRKWRERQRLQELDAASKVTEQREKAKDLLEWNQKQSOVKNNKINRASEA
FVKSEKETPGTEWKEVAQLCDENFPKSSKOCKDVSRILRSVLMSLKPTLSRL"

ORIGIN
Query Match 84.4%; Score 451.8; DB 11; Length 636;
Best Local Similarity 90.3%; Pred. No. 2e-92;
Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 1 CGGCCGCTTCTCGCCCGACGAGCGGAGATTGCTGGCATCGAAGTACCTCGGTT 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 CGGCCGCTTCTCGCCCGACGAGCGGAGATTGCTGGCATCGAAGTACCTCGGTT 130

```



```
/db_xref="taxon:32630"
/clone="FLH025389.01L"
/lab_host="Escherichia coli DH5alpha T1 resistant"
/note="derived from MGC template"
1. .>636
/gene="CLTB"
1. .>636
/gene="CLTB"
/note="Lcb"
/codon_start=1
/transl_table=11
/product="clathrin light polypeptide"
/protein_id="AA29302.1"
/db_xref="GI:60653215"
/translation="MADDFGFSSSESGAPEAABEDPAAAFIAQOESIEAGIENDEGF
GAPAGSHAAPAGPTSGAGSDMGTTVNGDFQANGPADGYAIAQADRLTQBPES
IRKWEQRKQLDLAASKVTEQWEKAKKDLSEWNQROSEQVEKKNINRASEEA
FVKESKETPGTEWEKVAQLCDFNPKSKQCKDVSRSLRSLVLSLKQTPLSRL"

ORIGIN
Query Match      84.4%; Score 451.8; DB 11; Length 636;
Best Local Similarity 90.3%; Pred. No. 2e-92;
Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1 CGGCGGCTTCTGCGCCAGCAGGAGCGAGATTGCTGGCATCGAGATCACTCGGGTT 60
DB 1 CGGCGGCTTCTGCGCCAGCAGGAGCGAGATTGCGAGCATAGAACGACGAGGCT 130
QY 61 TCGGGGCACTTGCCTGCGCCAGCAGGAGCGAGATTGCTGGCATCGAGATCACTCGGGTT 120
DB 131 TCGGGGCACTTGCCTGCGCCAGCAGGAGCGAGATTGCGAGCATAGAACGACGAGGCT 190
QY 121 GTTCGGAGCAATGGGACTACAGTCAATGGAGATGTGTTTCAGAGAGCTTAACGGGCTG 180
DB 191 GTTCTGAGCAATGGGAGCCACAGTCAATGGAGATGTGTTTCAGGAGGCCAACGGTCT 250
QY 181 CGATGGCTACGCTCGGATTGCCAGCGGAGCAGGTTGCTCAGCAGGCTGAGAGCATCC 240
DB 251 CTGATGGCTACGACGCCATTGCCAGGCTGACAGGCTGACCCAGGAGCTGAGAGCATCC 310
QY 241 GCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 311 GCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 370
QY 301 CCGAACAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 371 CCGAACAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 430
QY 361 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 431 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 490
QY 421 AATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 491 AATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 550
QY 481 ACCCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
DB 551 ACCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605

RESULT 10
BT007759
LOCUS      636 bp mRNA linear SYN 13-MAY-2003
DEFINITION Synthetic construct Homo sapiens clathrin, light polypeptide (Lcb)
            mRNA, partial cds.
ACCESSION  BT007759
VERSION    BT007759.1 GI:30584356
KEYWORDS  FLI CDNA.
SOURCE     synthetic construct
ORGANISM  other sequences; artificial sequences.
REFERENCE  1 (bases 1 to 636)

Kalinine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
Cloning of human full-length CDSs in BD Creator(TM) System Donor
vector
Unpublished
2 (bases 1 to 636)
Kalinine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
Direct Submission
Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
This CDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide Kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
Location/Qualifiers
1. .636
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH00332L1.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/note="Vector: pDNR-Dual"
1. .>636
/note="Mutations: 635:Stop->Leu"
/codon_start=1
/transl_table=11
/product="Homo sapiens clathrin, light polypeptide (Lcb)"
/protein_id="AAP36427.1"
/db_xref="GI:30584357"
/translation="MADDFGFSSSESGAPEAABEDPAAAFIAQOESIEAGIENDEGF
GAPAGSHAAPAGPTSGAGSDMGTTVNGDFQANGPADGYAIAQADRLTQBPES
IRKWEQRKQLDLAASKVTEQWEKAKKDLSEWNQROSEQVEKKNINRASEEA
FVKESKETPGTEWEKVAQLCDFNPKSKQCKDVSRSLRSLVLSLKQTPLSRL"

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .636
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH00332L1.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/note="Vector: pDNR-Dual"
1. .>636
/note="Mutations: 635:Stop->Leu"
/codon_start=1
/transl_table=11
/product="Homo sapiens clathrin, light polypeptide (Lcb)"
/protein_id="AAP36427.1"
/db_xref="GI:30584357"
/translation="MADDFGFSSSESGAPEAABEDPAAAFIAQOESIEAGIENDEGF
GAPAGSHAAPAGPTSGAGSDMGTTVNGDFQANGPADGYAIAQADRLTQBPES
IRKWEQRKQLDLAASKVTEQWEKAKKDLSEWNQROSEQVEKKNINRASEEA
FVKESKETPGTEWEKVAQLCDFNPKSKQCKDVSRSLRSLVLSLKQTPLSRL"

Query Match      84.4%; Score 451.8; DB 11; Length 636;
Best Local Similarity 90.3%; Pred. No. 2e-92;
Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1 CGGCGGCTTCTGCGCCAGCAGGAGCGAGATTGCTGGCATCGAGATCACTCGGGTT 60
DB 1 CGGCGGCTTCTGCGCCAGCAGGAGCGAGATTGCGAGCATAGAACGACGAGGCT 130
QY 61 TCGGGGCACTTGCCTGCGCCAGCAGGAGCGAGATTGCTGGCATCGAGATCACTCGGGTT 120
DB 131 TCGGGGCACTTGCCTGCGCCAGCAGGAGCGAGATTGCGAGCATAGAACGACGAGGCT 190
QY 121 GTTCGGAGCAATGGGACTACAGTCAATGGAGATGTGTTTCAGAGAGCTTAACGGGCTG 180
DB 191 GTTCTGAGCAATGGGAGCCACAGTCAATGGAGATGTGTTTCAGGAGGCCAACGGTCT 250
QY 181 CGATGGCTACGCTCGGATTGCCAGCGGAGCAGGTTGCTCAGCAGGCTGAGAGCATCC 240
DB 251 CTGATGGCTACGACGCCATTGCCAGGCTGACAGGCTGACCCAGGAGCTGAGAGCATCC 310
QY 241 GCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 311 GCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 370
QY 301 CCGAACAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 371 CCGAACAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 430
QY 361 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 431 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 490
QY 421 AATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 491 AATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 550
QY 481 ACCCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
DB 551 ACCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605

BT007759      636 bp mRNA linear SYN 13-MAY-2003
LOCUS      Synthetic construct Homo sapiens clathrin, light polypeptide (Lcb)
DEFINITION  mRNA, partial cds.
ACCESSION  BT007759
VERSION    BT007759.1 GI:30584356
KEYWORDS  FLI CDNA.
SOURCE     synthetic construct
ORGANISM  other sequences; artificial sequences.
REFERENCE  1 (bases 1 to 636)
```

```

Qy 361 GTGAACAGGTGAGAGAACAGATCAACAACAGGGGATCGGAAGAGCGCTTTGTGAAG 420
Db 431 GTGAACAGGTAGAGAGAACAGATCAACAACCGGGCATCCGAGAGGGCTTTGTGAAG 490
Qy 421 AATCCAAGGAGGAGACCCAGGACACAGTGGGAGAGGTGGCCAGCTGTGTGACTTCA 480
Db 491 AATCCAAGGAGGAGACCCAGGACACAGTGGGAGAGGTGGCCAGCTGTGTGACTTCA 550
Qy 481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTCGCTCGGTGCTCAT 535
Db 551 ACCCAAGAGCAGCAAGCAGTGCACAGATGTGTCCCGCTCGCTCGGTGCTCAT 605

RESULT 11
HUMLC2HB 1051 bp mRNA linear PRI 07-JAN-1995
LOCUS Human lymphocyte clathrin light-chain B mRNA, complete cds.
ACCESSION M20470.1 GI:187056
VERSION M20470.1 GI:187056
KEYWORDS clathrin; clathrin light chain b.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1051)
Jackson, A.P. and Parham, P.
Structure of human clathrin light chains. Conservation of light
chain polymorphism in three mammalian species
J. Biol. Chem. 263 (32), 16688-16695 (1988)
3267234
ORIGINAL SOURCE text: Human lymphocyte (cell line T7527), cDNA to
mRNA.
Draft entry and computer-readable sequence [1] kindly submitted by
A. Jackson, 02-SEP-1988.
FEATURES
source
Location/Qualifiers
1..1051
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="Unassigned"
1..1051
/gene="CLTB"
140..775
/gene="CLTB"
/notes="clathrin light-chain B"
/codon_start=1
/protein_id="AA59506.1"
/db_xref="GI:307119"
/db_xref="GDB:G00-125-861"
/translation="MADDFGFFSSGSAPEAREDPAAFLAQSEIAGIENDEGF
GAPGSHAAPQAPQPTSGAGSDMGTTVNGDVFQANGPADYAAIAQADRLTQEPES
IRKWRERQKELQSLDAASKVTEQEWREKAKDLEWNQROSEOVENKKNRASEA
FVKESKETPTETHEKVALQCDLPFKSKQCKDVSRLRSVLMSLKQPLSR"
ORIGIN 33 bp upstream of Hpall site.
Query Match 84.4%; Score 451.8; DB 8; Length 1051;
Best Local Similarity 90.3%; Pred. No. 2.1e-92;
Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
Qy 1 CGGCGCGCTTCTCGGCCAGAGAGCGAGATGCTGGCATCGAGATGACTCGGGTT 60
Db 210 CGGCGCGCTTCTCGGCCAGAGAGCGAGATGCTGGCATCGAGATGACTCGGGCT 269
Qy 61 TCGGGGACCTGCGCCAGCAGCGAGTGGCTCTGCGCAGCCCGGACTCGCAGCGGGGTG 120
Db 270 TCGGGGACCTGCGCGAGCGCATGCGGCCCGCCCGCGAGCGGGCCACGAGTGGGCTG 329
Qy 121 GTTCGGAGGACATGGGGACTACAGTCAATGAGAGATGTGTTTTCAGGAGCTTAACGGGCTG 180
Db 330 GTTCTGAGGACATGGGGACCAAGTCAATGAGAGATGTGTTTTCAGGAGGCCAACGGTCTG 389

```

```

Qy 181 CCGATGGCTACGCTGCGATTGTCAGCGCGACAGGTGTAAGTCTCAGGAGCGCTGAGAGCATCC 240
Db 390 CTGATGGCTAGCAGCCATTGCCAGGCTGACAGGCTGACCCAGGAGCGCTGAGAGCATCC 449
Qy 241 GCAAGTGGAGAGGAGGAGCAGAGAAAGAGGTGTCAGGAGTTGGATGTCCTCGAAGGTGA 300
Db 450 GCAAGTGGCGAGAGGAGCAGAGAAACGGCTGCAAGAGCTGGATGCTGCATCTAAGGTCA 509
Qy 301 CGGAACAGGAGTGGCGGAGAGCGCCAAAGAACCTGGAGGAGTGAACAGCGCCCAA 360
Db 510 CGGAACAGGAAATGGCGGAGAGCGCCAAAGAGGACCTGGAGGAGTGAACAGCGCCAGA 569
Qy 361 GTGAACAGGTTGAGAGAAACAAGATCAACAACAGGGCATCGGAAGAGCGCTTTTGTGAAG 420
Db 570 GTGAACAGTAGAGAGAAACAAGATCAACAACCGGGCATCCGAGAGCGCTTTCTGGAAG 629
Qy 421 AATCCAAGGAGGAGACCCAGGACACAGAGTGGGAGAAAGGTGGCCAGCTGTGTGACTTCA 480
Db 630 AATCCAAGGAGGAGACCCAGGACACAGAGTGGGAGAAAGGTGGCCAGCTGTGTGACTTCA 689
Qy 481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTCGCTCGGTGCTCAT 535
Db 690 ACCCAAGAGCAGCAAGCAGTGCACAGATGTGTCCCGCTCGCTCGGTGCTCAT 744

RESULT 12
BC006457 1099 bp mRNA linear PRI 08-MAR-2005
LOCUS Homo sapiens clathrin, light polypeptide (lcb), transcript variant
DEFINITION nonbrain, mRNA (cdna clone MGC:1776 IMAGE:3538261), complete cds.
ACCESSION BC006457
KEYWORDS MGC.
SOURCE BC006457.1 GI:13623662
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1099)
Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD,
Collins FS, Wagner L, Shenmen CM, Schuler GD, Altshul SF, Zeeberg
B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T,
Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin
GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL,
Scheetz TE, Brownstein MJ, Usdin TB, Toshiyuki S, Carninci P,
Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy
SJ, Bosak SA, McEwan EJ, McKernan KJ, Malek JA, Gunaratne PH,
Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW,
Villalón DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J,
Halton E, Kettelman M, Madan A, Rodriguez S, Sanchez A, Whiting M,
Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW,
Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J,
Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U,
Smailus DB, Schnerch A, Schein JE, Jones SJ and Marra MA.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1099)
Director MGC Project.
Direct Submission
TITLE Submitted (09-APR-2001) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NTH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/BTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,

```


BC Cancer Agency, Vancouver, BC, Canada
 info@bcgsc.bc.ca
 Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
 Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
 Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
 Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
 Teika Olson, Diana Palquist, Anca Petrescu, Anna Liisa Prabhu,
 Parvaneh Saedi, JR Santos, Angélique Schnerch, Ursula Skaleka,
 Duane Smalhus, Jeff Stott, Miranda Tsai, George Yang, Jacque
 Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 8 Row: j Column: 21
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 32483393.

FEATURES

source

1..1099
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:1776 IMAGE:3538261"
 /tissue type="lung, small cell carcinoma"
 /clone_lib="NIH MGC 7"
 /lab host="DH10B-R"
 /note="Vector: pOTB7"

gene

1..1099
 /gene="CLTB"
 /note="synonym: LCB"
 /db_xref="GeneID:1212"
 /db_xref="MIM:118970"
 130..765
 /gene="CLTB"
 /codon_start=1
 /product="clathrin, light polypeptide, isoform a"

CDS

/protein_id="AAH06457.1"
 /db_xref="GI:13621663"
 /db_xref="GeneID:1212"
 /db_xref="MIM:118970"
 /translation="MADDFGFSSSEGAPEAAEDPAAFLAQESFIAGIENDEGF
 GAPGSHAQRKLPQGTSGSDMTGVNDFVQGNPAGDGYAAIAQADRLTORPES
 IAKWEORKRLOELDAASKVTEQWREKAKDLBNORQSEQVEKRNKNRASEEA
 FVKSEKPTGTEWEKVAQLCDNFPSKQCKDVSRLSLVLSLQTPLSR"

ORIGIN

Query Match 84.4%; Score 451.8; DB 8; Length 1099;
 Best Local Similarity 90.3%; Pred. No. 2.1e-92;
 Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 QY 1 CGGCGCTTCTTGGCCGACGAGGAGCGAGATTGCTGGCATCGAGATGACTCGGTT 60
 DB 200 CGGCGCTTCTTGGCCGACGAGGAGCGAGATTGCGAGGAGGAGGAGGAGGCT 259
 QY 61 TCGGGACCTCGCGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
 DB 260 TCGGGACCTCGCGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 319
 QY 121 GTTCGAGGACATGGGAGCTACAGTCAATGGAGATGTTTTCAGGAGGCTAACGGGCTG 180
 DB 320 GTTCTGAGGACATGGGAGCCACAGTCAATGGAGATGTTTTCAGGAGGCAACGGGCTG 379
 QY 181 CGATGGCTACGCTGGATTTGCCAGCGGACAGGTTGACTCAGGAGCTGAGAGCTCC 240
 DB 380 CTGATGGCTACGAGCCATTGGCCAGGCTGACAGGCTGACCCAGGAGGCTGAGAGCTCC 439
 QY 241 GCAAGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
 DB 440 GCAAGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 499
 QY 301 CGGAACAGGAGTGGCGGAGAGAGGCCAAAGAACCTGGAGGAGTGGAACCGAGCGCA 360
 DB 500 CGGAACAGGAGTGGCGGAGAGAGGCCCAAGAGGAGGAGGAGGAGGAGGAGGAGG 559

QY 361 GTGAACAGGTTGAGAAGAACAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAG 420
 DB 560 GTGAACAGTAGAGAGAGAACAGATCAACAACCGGGCATCCGAGGAGGCTTTCTGTGAAG 619
 QY 421 AATCAAGGAGGAGACCCAGGACAGAGTGGGAGAGGAGGAGGAGGAGGAGGAGGAG 480
 DB 620 AATCAAGGAGGAGACCCAGGACAGAGTGGGAGAGGAGGAGGAGGAGGAGGAGGAG 679
 QY 481 ACCCTAAGGAGGAGCAAGCAATGTAAGACGTGTCCCGCTCGGCTCGGCTCGGCTCAT 535
 DB 680 ACCCAAGAGCAGCAGCAGCAGTGTCCCGCTCGGCTCGGCTCGGCTCAT 734

RESULT 13

BC006332

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBLISHED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC006332 1164 bp mRNA linear PRI 08-MAR-2005
 Homo sapiens clathrin, light polypeptide (Lcb), transcript variant
 nonbrain, mRNA (cdna clone MGC:12930 IMAGE:4299637), complete cds.

BC006332
 BC006332.2 GI:33991210

MGC.

Homo sapiens (human)

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Homo.

1 (bases 1 to 1164)

Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD,
 Collins FS, Wagner L, Sherman CM, Schuler GP, Altshul SF, Zeeberg
 B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T,
 Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin
 GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL,
 Scheetz TE, Brownstein MJ, Ustin TB, Toshiyuki S, Carninci P,
 Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullaly
 SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH,
 Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW,
 Villalón DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J,
 Helton E, Kettman M, Madan A, Rodriguez S, Sanchez A, Whiting M,
 Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW,
 Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J,
 Schmutz J, Myers RM, Butterfield J, Krzywinski MI, Skalska U,
 Smalusz DB, Schnerch A, Schein JE, Jones SJ and Marra MA.
 Mammalian Gene Collection Program Team
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 1164)

Director MGC Project.

Direct Submission

Submitted (09-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 On Aug 20, 2003 this sequence version replaced gi:13623458.
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov

Blakesley R.W., Bouffard G.G., Beckstrom-Sternberg S.M., Benjamin B.,
 Akhter N., Ayale K., Green K., Brinkley C., Brooks S.,
 Dietrich N.L., Granite S., Guan X., Gupta J., Haghighi P.,
 Hansen N., Ho S.-L., Karlins E., Kwong P., Laric P., Legaspi R.,
 Maduro Q.L., Mastello C., Maskeri B., Mastrian S.D., McCloskey J.C.,
 McDowell J., Pearson R., Stantripop S., Thomas P.J., Touchman J.W.,
 Tsugeon C., Vogt J.L., Walker M.A., Wetherby K.D., Wiggins L.,

Young, A.: Zhang, L.-H. and Green, E. D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 17 Row: m Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction.

FEATURES

Location/Qualifiers
1. 1164

1. .1164

NOTT: T

```
/organism="Homo
```

/mol_type="mRNA"

/db_xref="taxon:9606"

```
/clone="MGC:12930 IMAGE:4299637"
```

/tissue type="Eve. normal

```
/clone Lib="NTH MGC 43"
```

UD-80 LNU"-+BOq qeL/
COM ITN =ATT_ayroT/

gene

1. .1164

```

/ gene = "CI.TB"

```

```
/genre= CIDB
/notes=Symptoms: T.CB"
```

[/post=»выполучим: псв»](#)
[/ak.html=»выполучим: псв»](#)

```
/db_xref="GeneID:1212"
```

```
/db_xref="MIM:118970"
```

144. .779

/ gene = "CLTB"

/codon start=1

```

/product="clathrin. light polymeptide. isoform a"

```

[http://www.fishbase.org](#)

/p/role11_id=AAH06332.1
 /q/ab=904.12633450"

/db_xref="GI:13623459"

/db_xref="GeneID:1212"

```
/db_xref="MIM:118970"
/translators="MADDFG7SSSEGAPEAEEDPAAAFILAQOESEIAGIENDEGF
GAPAGSHAAPQAPGPTSGAGSDMGTTGVDFQZANGPAGYAAIAQDLKTQPE
IRKWEQEKRLQGLDASKVTEQEMKAKCLLEWNQOQSEVKNKNNRABEA
FVPEKKEETPTGWEKVAOICDNPKKSKCKDVSRVUTMSIKOTPTI"

```

ORIGIN

Query Match 84.4%; Score 451.8; DB 8; Length 1164;

Query Macchi	84.4%;	Score 431.8;	DB 8;	Length 1104;
Best Local Similarity	90.3%;	Pred. No. 2.1e-92;		

Best local similarity 90.3%; PRed: NO. 2.1e-92;
Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy	1	CGCGCGCTTCTTGGCCCGACAGAGAGCAGAGATTGCTGGCANTCGAAGATGACTCGGGTT	60
Db	214	CGCGCGCGCTTCTTGGCCCGACAGAGAGCAGAGATTGAGGCATAGAGAAACGACGAGGGCT	273
Qy	61	TCGGGGCACCTTCGCGGCAGCAGCGTGGCTCTGCGCAGCCCGGACTCGCGAGCGGGGGTG	120
Db	274	TCGGGGCACCTTCGCGGCAGCAGCGTGGCTCCCGCGCAGCCCGGGCCCCACGATGGGGCTG	333
Qy	121	GTTCGGAGGACATGGGGACTPACAGTCAATCGAGATGTGTTTCAAGGAGGCTTAAACGGGCTTG	180
Db	334	GTTCGAGGACATGGGGACACAGTCAATGAGAGATGTGTTTCAAGGAGGCCAAACGGTCTTG	393
Qy	181	CCGATGGCTACGCTGCGATTGCCAGCGCAGTAGTTGACTCAGGAGGCTTCAGAGCATCC	240
Db	394	CTGATGCTACGAGAGCATTTGCCAGGCTGACAGGCTGACCCAGGAGGCTTCGAGAGCATCC	453
Qy	241	GCAAGTGGAGAGGAGGACAGAAAGGCTGCAGGAGTTGGATGCTGCTCCCTCGAAGGTGA	300
Db	454	GCAAGTGGCGAGAGGAGCAGAGGAAACGGCTGCAAGAGCTGGATGCTGCATCTAAGTCA	513
Qy	301	CCGAAACAGGAGTGGCGGGAGAGCCCAAAAAGACTTGGAGGAGTGGAAACAGGCGCCAAA	360
Db	514	CCGAAACAGGAAATGGCGGGAGAGCCCAAGAGGACCTTGGAGGAGTGGAAACAGGCGCAGA	573
Qy	361	GTGAAACAGGTTTGAGAGCAACAGATCAACAAAGGGCATCGGAAGAGGCTTTTGTGTAAG	420
Db	574	GTGAAACAGTAGAGAGAAACAGATCAACAAACCGGGCATCCGAGGAGGCTTTTGTGTAAG	633
Qy	421	AATCCAAGGAGGAGACCCACGGCACAAGATGGGAGAAAGTGGGCCACGCTGTGTGACTTCA	480
Db	634	AATCCAAGGAGGAGACCCACGGCACAAGTGGGAGAAAGTGGGCCACGCTGTGTGACTTCA	693

Db	630	AATCCAAGGAGAGACCCCGGCGCACAGAGTGGGGAAGGTGGCCGAGCTGTGTGACTTCA	689
Qy	481	ACCCTAAGCAGCAAGCAATTAAGACGTTGCCGCGCTGGCTCGGTGTCAT	535
Db	690	ACCCCAAGCAGCAAGCAGTGCAAAGACGTTGCCGCTGGCTCGGTGTCAT	744

Search completed: January 5, 2006, 20:29:50
Job time : 2906 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 5, 2006, 17:19:45 ; Search time 487 Seconds
(without alignments)
7321.582 Million cell updates/sec

Title: US-10-664-705-145
Perfect score: 535
Sequence: 1 cggccgcttctgcccag.....cgctgcgtcggtgctcat 535

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 21:.*
1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002as.*
7: geneseqn2002bs.*
8: geneseqn2003as.*
9: geneseqn2003bs.*
10: geneseqn2003cs.*
11: geneseqn2003ds.*
12: geneseqn2004as.*
13: geneseqn2004bs.*
14: geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	535	100.0	535	ADR15112	ADR15112 Rat elect
2	471	88.0	982	ABK63500	ABK63500 Rat seque
3	471	88.0	982	ABK58810	ABK58810 Toxicity-
4	471	88.0	982	ABK53521	ABK53521 Primary r
5	471	88.0	982	ABT42334	ABT42334 Toxicity
6	471	88.0	982	ADW21848	ADW21848 Rat hepat
7	471	88.0	982	ADP72639	ADP72639 Renal tox
8	471	88.0	982	ADV41572	ADV41572 Rat cardi
9	451.8	84.4	1051	ABK83950	ABK83950 Human cDN
10	451.8	84.4	1051	ADJ62744	ADJ62744 Human cDN
11	451.8	84.4	1051	ADK07445	ADK07445 Cyclin-de
12	387.8	72.5	1134	ADK07447	ADK07447 Cyclin-de
13	380.8	71.2	894	ABK34444	ABK34444 Human cDN
14	379.2	70.9	907	ABK34409	ABK34409 Human cDN
15	364.4	68.1	490	ACH32890	ACH32890 Human end
16	293.8	54.9	493	ACH39643	ACH39643 Human foe
17	268.8	50.2	485	ACH40125	ACH40125 Human foe
18	240.2	44.9	562	ADQ52771	ADQ52771 Novel can
19	213.4	39.9	853	ADP04090	ADP04090 Human col

20	213.4	39.9	1023	13	ACN38893	Acn38893 Tumour-as
21	213.4	39.9	1164	6	ABQ54448	ABQ54448 Human ova
22	213.4	39.9	1233	12	ADP04094	Adp04094 Human col
23	130.8	35.7	1262	12	ADP04095	Adp04095 Human col
24	176.6	33.0	530	14	ACL55096	ACL55096 Human col
25	176.2	32.9	958	6	ABI99595	Abi99595 Mouse lac
26	175	32.7	528	14	ACL62179	ACL62179 Human col
27	175	32.7	1666	12	ADP04091	Adp04091 Human col
28	145.4	27.2	1958	12	ADP04093	Adp04093 Human col
29	143.4	26.8	1124	10	ADB53376	Adb53376 Primary r
30	143.4	26.8	1124	13	ADVA1478	Adv41478 Rat cardi
31	139.4	26.1	527	6	ABI99585	Abi99585 Mouse lac
32	138.6	25.9	1096	13	ADQ87410	Adq87410 Human tum
33	138.6	25.9	1105	6	ABK83951	ABK83951 Human cDN
34	138.6	25.9	1105	6	ABN95791	ABn95791 Gene #228
35	138.6	25.9	1105	13	ACN38894	Acn38894 Tumour-as
36	137.2	25.6	459	9	ACH40991	Ach40991 Human foe
37	130.8	24.4	5938	6	ABK84688	ABk84688 Human cDN
38	124.6	23.3	1542	12	ADP04092	Adp04092 Human col
39	121	22.6	735	10	ADC39087	Adc39087 Novel hum
40	115.4	21.6	473	9	ACH22936	Ach22936 Human adu
41	87.4	16.3	417	10	ADG10314	Adg10314 Mouse neu
42	78.2	14.6	1084	4	ABL29421	Abi29421 Drosophil
43	78.2	14.6	3201	4	ABL29420	Abi29420 Drosophil
44	72.6	13.6	484	9	ACH33747	ACH33747 Human end
45	72.6	13.6	617	13	ADU11758	Adu11758 Solid tum

ALIGNMENTS

RESULT 1
ADR15112
ID ADR15112 standard; DNA; 535 BP.
XX
AC ADR15112;
XX
DT 21-OCT-2004 (first entry)
XX
DE Rat electroconvulsive seizure (EC) signature gene seqid 145.
XX
KW neuroleptic; nootropic; antidepressant; eating disorder; tranquilizer;
KW neuropsychiatric disorder; signature gene; electroconvulsive seizure;
KW ECS; schizophrenia; autism; major depressive disorder; MDD;
KW bipolar affective disorder; BAD; psychotic depression; anxiety disorder;
KW eating disorder; attention deficit hyperactivity disorder;
KW neuropsychiatric drug; gene; ss; rat.
XX
OS Rattus norvegicus.
XX
PN US2004152107-A1.
XX
PD 05-AUG-2004.
XX
PF 18-SEP-2003; 2003US-00664705.
PR 18-SEP-2002; 2002US-0411718P.
PR 09-DEC-2002; 2002US-0431882P.
PR 18-JUN-2003; 2003US-0479970P.
PA (ALTA/) ALTAR C A.
PA (LAEN/) LAENG P.
PA (YOUN/) YOUNG T A.
PA (CHAR/) CHARLES V D.
PI Altar CA, Laeng P, Young TA, Charles VD;
XX WPI; 2004-580183/56.
DR Identifying compound to treat neuropsychiatric disorder, by contacting
PT cell with test compound, determining expression of signature genes by
PT cell, comparing expression of signature genes to expression in cell not
PT contacted with compound.

CC gene expression information may be used as diagnostic markers for the
CC prediction or identification of the physiological state of tissue or cell
CC sample that has been exposed to a compound or agent. Hepatotoxicity is
CC characterised by centrilobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent
XX
XX

SQ Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;

Query Match 88.0%; Score 471; DB 6; Length 982;
Best Local Similarity 90.8%; Pred. No. 1.5e-115;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 CGGCGCGCTTCTGCGCCAGCAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 60
DB 99 CGGCGCGCTTCTGCGCCAGCAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 158
QY 61 TCGGGGCACTTCCGCGCCAGCAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTTG 120
DB 159 TCGGGGCACTTCCGCGCCAGCAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTTG 218
QY 121 GTTCGGAGGACATGCGGACTACAGTCAATGGAGATGTTTTCAGGAGGCTAACGGGCTG 180
DB 219 GTTCGGAGGACATGCGGACTACAGTCAATGGAGATGTTTTCAGGAGGCTAACGGGCTG 278
QY 181 CCGATGGCTACGCTCGGACTACAGTCAATGGAGATGTTTTCAGGAGGCTAACGGGCTG 240
DB 279 CCGATGGCTACGCTCGGACTACAGTCAATGGAGATGTTTTCAGGAGGCTAACGGGCTG 338
QY 241 GCAAGTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
DB 339 GCAAGTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 398
QY 301 CCGAACAGGAGTGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
DB 399 CCGAACAGGAGTGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 458
QY 361 GTGAACAGGTTGAGAAGAACAGATCAACAGGATCGCTGACAAAGGCTTCTACCGAGC 394
DB 459 GTGAACAGGTTGAGAAGAACAGATCAACAGGATCGCTGACAAAGGCTTCTACCGAGC 518
QY 395 -----GCCATCGGAAGAGGCTTTTGTGAAGAGATCCA 426
DB 519 AGCCAGATCGTATACCATGGCTATGTCGCTATCGGATCGGAGGCTTTTGTGAAGAGATCCA 578
QY 427 AGGAGGAGACCCAGGACAGAGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 486
DB 579 AGGAGGAGACCCAGGACAGAGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 638
QY 487 AGAGCAGCAAGCAATGTAAGAGCTGTCCCGCTCGGCTCGGCTGCTCAT 535
DB 639 AGAGCAGCAAGCAATGTAAGAGCTGTCCCGCTCGGCTCGGCTGCTCAT 687

RESULT 3

ID ADB58810 standard; DNA; 982 BP.

AC ADB58810;

XX 04-DEC-2003 (first entry)

XX Toxicity-related gene, SEQ ID 3836.

DE Toxic; toxin; gene expression profile; hepatotoxicity; liver;
KW drug screening; toxicity assay; ds.
XX

OS Unidentified.

XX W02003064624-A2.

XX 07-AUG-2003.

PD

PF 31-JAN-2003; 2003WO-US003194.
XX
XX 31-JAN-2002; 2002US-00060087.
PR 15-MAR-2002; 2002US-0364045P.
PR 15-MAR-2002; 2002US-0364055P.
PR 30-DEC-2002; 2002US-0436643P.
XX
XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

XX WPI; 2003-689530/65.

XX Predicting a toxic effect of a compound, useful in identifying toxicity
PT markers in liver tissues or cells for drug screening and toxicity assays,
PT comprises preparing gene expression profile of tissue or cells exposed to
PT the compound.

XX Claim 1; SEQ ID NO 3836; 1156pp; English.

XX The present invention relates to a method for predicting a toxic effect
CC of a compound. The method comprises preparing a gene expression profile
CC of a tissue or cell sample exposed to the compound, and comparing the
CC gene expression profile to a database comprising SEQ ID 1-4925, where
CC differential expression of the gene indicates at least one toxic effect.
CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in
CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;

Query Match 88.0%; Score 471; DB 10; Length 982;

Best Local Similarity 90.8%; Pred. No. 1.5e-115;

Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 CGGCGCGCTTCTGCGCCAGCAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 60
DB 99 CGGCGCGCTTCTGCGCCAGCAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 158
QY 61 TCGGGGCACTTCCGCGCCAGCAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTTG 120
DB 159 TCGGGGCACTTCCGCGCCAGCAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTTG 218
QY 121 GTTCGGAGGACATGCGGACTACAGTCAATGGAGATGTTTTCAGGAGGCTAACGGGCTG 180
DB 219 GTTCGGAGGACATGCGGACTACAGTCAATGGAGATGTTTTCAGGAGGCTAACGGGCTG 278
QY 181 CCGATGGCTACGCTCGGACTACAGTCAATGGAGATGTTTTCAGGAGGCTAACGGGCTG 240
DB 279 CCGATGGCTACGCTCGGACTACAGTCAATGGAGATGTTTTCAGGAGGCTAACGGGCTG 338
QY 241 GCAAGTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
DB 339 GCAAGTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 398
QY 301 CCGAACAGGAGTGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
DB 399 CCGAACAGGAGTGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 458
QY 361 GTGAACAGGTTGAGAAGAACAGATCAACAGGATCGCTGACAAAGGCTTCTACCGAGC 394
DB 459 GTGAACAGGTTGAGAAGAACAGATCAACAGGATCGCTGACAAAGGCTTCTACCGAGC 518
QY 395 -----GCCATCGGAAGAGGCTTTTGTGAAGAGATCCA 426

Db 519 AGCCAGATGCTGATACCAATGGCTATGTGGCATCGGAAGAGGCTTTGTGAAAGATCCA 578
 Qy 427 AGGAGGAGACCCAGGACACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAACCCCTA 486
 Db 579 AGGAGGAGACCCAGGACACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAACCCCTA 638
 Qy 487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTGGCGTGGTGTCTCAT 535
 Db 639 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTGGCGTGGTGTCTCAT 687

RESULT 4
 ADB53521
 ID ADB53521 standard; DNA; 982 BP.
 AC ADB53521;
 DT 04-DEC-2003 (first entry)
 XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4063.
 DE toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
 KW toxicity marker; toxicity progression; drug screening;
 KW primary rat hepatocyte toxicity modelling; gene; ds.
 XX Rattus norvegicus.
 OS
 XX
 XX WO2003065993-A2.
 XX
 XX 14-AUG-2003.
 XX
 XX 04-FEB-2003; 2003WO-US003482.
 XX
 XX 04-FEB-2002; 2002US-0353171P.
 PR 13-MAR-2002; 2002US-0363534P.
 PR 08-APR-2002; 2002US-0370248P.
 PR 10-APR-2002; 2002US-0371134P.
 PR 10-APR-2002; 2002US-0371135P.
 PR 10-APR-2002; 2002US-0371150P.
 PR 11-APR-2002; 2002US-0371413P.
 PR 19-APR-2002; 2002US-0373601P.
 PR 19-APR-2002; 2002US-0373602P.
 PR 22-APR-2002; 2002US-0374139P.
 PR 08-MAY-2002; 2002US-0378370P.
 PR 09-MAY-2002; 2002US-0378652P.
 PR 09-MAY-2002; 2002US-0378653P.
 PR 09-MAY-2002; 2002US-0378665P.
 PR 09-JUL-2002; 2002US-0394230P.
 PR 09-JUL-2002; 2002US-0394253P.
 PR 04-SEP-2002; 2002US-0407688P.
 PR 28-JAN-2003; 2003US-0442900P.
 XX (GENE-) GENE LOGIC INC.
 XX
 XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
 PI Elashoff M;
 XX
 XX WPI; 2003-731472/69.
 XX
 XX Determining if a compound induces a toxic effect on a tissue or cell, for
 PT identifying hepatotoxic compounds, comprises comparing a gene expression
 PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
 PT mean values.
 XX
 XX Claim 44; SEQ ID NO 4063; 874pp; English.
 XX
 XX The present invention describes a method for determining whether a
 CC compound induces a toxic effect on a tissue or cell. The method comprises
 CC preparing a gene expression profile of a tissue or cell sample exposed to
 CC the compound, and comparing the gene expression profile to a database
 CC comprising data or information on the Tox mean and non-Tox mean value.
 CC The method is useful for predicting or identifying at least one toxic

CC effect, particularly hepatotoxicity, of a test or unknown compound. The
 CC genes listed in the specification are useful as diagnostic or toxicity
 CC markers for the prediction or identification of the physiological state
 CC of tissue or cell sample that has been exposed to a compound, or to
 CC identify or predict the toxic effects of a compound or an agent. These
 CC may also be used as markers for monitoring toxicity progression or for
 CC drug screening. The present sequence represents a primary rat hepatocyte
 CC toxicity modelling related gene sequence from the present invention.
 XX
 XX Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;
 SQ
 Query Match 88.0%; Score 471; DB 10; Length 982;
 Best Local Similarity 90.8%; Pred. No. 1.5e-115;
 Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;
 Qy 1 CGCGCGCTTCTCTGGCCACGAGAGCGAGATGCTGGCATCGAGAATGACTCGGGTT 60
 Db 99 CGCGCGCTTCTCTGGCCACGAGAGCGAGATGCTGGCATCGAGAATGACTCGGGTT 158
 Qy 61 TCGGGGCACCTGCGCCGAGCGAGTGGCTTGGCGAGCCGCGACTCGGAGCGGGGTG 120
 Db 159 TCGGGGCACCTGCGCCGAGCGAGTGGCTTGGCGAGCCGCGACTCGGAGCGGGGTG 218
 Qy 121 GTTCGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAAACGGGCTG 180
 Db 219 GTTCGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAAACGGGCTG 278
 Qy 181 CCGATGGCTACGCTGCGATTGCCAGCGGACAGTGTGACTCAGGAGCCTGAGAGCATCC 240
 Db 279 CCGATGGCTACGCTGCGATTGCCAGCGGACAGTGTGACTCAGGAGCCTGAGAGCATCC 338
 Qy 241 GCAAGTGGAGAGGAGCAGAGAAAGAGCTGCGAGGATGGATGCTGCTCGAAGTGA 300
 Db 339 GCAAGTGGAGAGGAGCAGAGAAAGAGCTGCGAGGATGGATGCTGCTCGAAGTGA 398
 Qy 301 CCGAACAGGAGTGGCGGAGAGGCGCAAAAGAGCTGGAGGAGTGAACCCAGCGCCAAA 360
 Db 399 CCGAACAGGAGTGGCGGAGAGGCGCAAAAGAGCTGGAGGAGTGAACCCAGCGCCAAA 458
 Qy 361 GTGAACAGGTTGAGAGAAACAAGATCAACAACAG----- 394
 Db 459 GTGAACAGGTTGAGAGAAACAAGATCAACAACAGTCCGACAAAGCGTTCTACCAAGC 518
 Qy 395 -----GGCATCGGAGAGGCTTTTGTGAAAGATCCA 426
 Db 519 AGCCAGATGCTGATACATTTGGCTATGTGGCATCGGAAGAGGCTTTTGTGAAAGATCCA 578
 Qy 427 AGGAGGAGACCCAGGACACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAACCCCTA 486
 Db 579 AGGAGGAGACCCAGGACACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAACCCCTA 638
 Qy 487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTGGCGTGGTGTCTCAT 535
 Db 639 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTGGCGTGGTGTCTCAT 687
 RESULT 5
 ADB42334
 ID ADB42334 standard; DNA; 982 BP.
 AC ADB42334;
 XX
 XX 26-JUN-2003 (first entry)
 DT
 XX Toxicity modelling related rat gene SEQ ID No 2036.
 XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;
 KW database; drug screening; toxicity assay; rat; ds.
 XX Rattus norvegicus.
 XX
 XX WO200295000-A2.
 XX
 XX

PD 28-NOV-2002.
XX 22-MAY-2002; 2002WO-US016173.
XX 22-MAY-2001; 2001US-0292335P.
XX 13-JUN-2001; 2001US-0297523P.
XX 19-JUN-2001; 2001US-0298925P.
XX 10-JUL-2001; 2001US-0303807P.
XX 10-JUL-2001; 2001US-0303808P.
XX 10-JUL-2001; 2001US-0303810P.
XX 28-AUG-2001; 2001US-0315047P.
XX 27-SEP-2001; 2001US-0324928P.
XX 22-OCT-2001; 2001US-0330462P.
XX 01-NOV-2001; 2001US-0330867P.
XX 21-NOV-2001; 2001US-0331805P.
XX 06-DEC-2001; 2001US-0336144P.
XX 19-DEC-2001; 2001US-0340873P.
XX 21-FEB-2002; 2002US-0357842P.
XX 21-FEB-2002; 2002US-0357843P.
XX 21-FEB-2002; 2002US-0357844P.
XX 15-MAR-2002; 2002US-0364134P.
XX 08-APR-2002; 2002US-0370144P.
XX 08-APR-2002; 2002US-0370206P.
XX 08-APR-2002; 2002US-0370247P.
XX 17-APR-2002; 2002US-0372794P.
XX 21-APR-2002; 2002US-0371679P.
XX (GENE-) GENE LOGIC INC.
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX WPI; 2003-148464/14.
XX Predicting at least one toxic effect of a compound, useful for toxicity
XX modeling, comprises preparing a gene expression profile of a tissue or
XX cell sample exposed to the compound, and comparing the gene expression
XX profile to a database.
XX Example 4; Page; 446pp; English.
XX The invention relates to a novel method of predicting at least one toxic
XX effect of a compound. The method comprises a gene expression profile of a
XX tissue or cell sample exposed to the compound, and comparing the gene
XX expression profile to a database comprising at least part of the data or
XX information given in the specification. The methods are useful for
XX predicting at least one toxic effect of a compound, predicting the renal
XX progression of a toxic effect of a compound, predicting the renal
XX toxicity of a compound, or identifying toxicity markers in tissues or
XX cells exposed to known renal toxin. The genes are useful as toxicity
XX markers in drug screening and toxicity assays, in monitoring disease or
XX physiological states, or disease progression. This polynucleotide
XX represents a rat DNA sequence relating to the toxic effect database
XX described in the specification. NOTE: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from the World Intellectual Property
XX Organization
XX Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;
Query Match 88.0%; Score 471; DB 10; Length 982;
Best Local Similarity 90.8%; Pred. No. 1.5e-115;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;
QY 1 CGGCGCCCTTCCCTGGCCAGCAGGAGGAGATTGCTGGCATCGAGATGATCGGGTT 60
DB 99 CGGCGCCCTTCCCTGGCCAGCAGGAGGAGATTGCTGGCATCGAGATGATCGGGTT 158
QY 61 TCGGGGACCTGCGCCAGCCAGGTCCTCTCGGCAGCCCGACTCGCGAGCGGGGTG 120
DB 159 TCGGGGACCTGCGCCAGCCAGGTCCTCTCGGCAGCCCGACTCGCGAGCGGGGTG 218
QY 121 GTTCGAGGACATGGGGACTACAGTCAATCGAGATGTTTTCAGGAGGCTTACGGCCCTG 180

Db 219 GTTCGAGGACATGGGGACTACAGTCAATCGAGATGTTTTCAGGAGGCTTACGGGCCCTG 278
QY 181 CCGATGGCTAGCTGCGGATTTGCCAGCGGACAGGTTGACTCAGAGAGCTTCAGAGCATCC 240
Db 279 CCGATGGCTAGCTGCGGATTTGCCAGCGGACAGGTTGACTCAGAGAGCTTCAGAGCATCC 338
QY 241 GCAAAGTGGAGAGAGGAGCAGAAAGAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
Db 339 GCAAAGTGGAGAGAGGAGCAGAAAGAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 398
QY 301 CCGAACAGGAGTGGCGGAGAGGCGCAAAAGAGGCTGCAGGAGTTGGAGGAGTGGAGGCGCCAAA 360
Db 399 CCGAACAGGAGTGGCGGAGAGGCGCAAAAGAGGCTGCAGGAGTTGGAGGAGTGGAGGCGCCAAA 458
QY 361 GTGAACAGGTTGAGAGAAAGAAAGATCAACACAG----- 394
Db 459 GTGAACAGGTTGAGAGAAAGAAAGATCAACACAGGATCGTGACAAAGCGTTCTACCAGC 518
QY 395 -----GGATCGAAGAGGCTTTTGTGAAAGATCCA 426
Db 519 AGCCAGATGCTGATACCATTTGGCTATGTGGCATCGAAGAGGCTTTTGTGAAAGATCCA 578
QY 427 AGGAGGAGACCCAGGAGCAGAGTGGGAGAGGTTGGCCAGCTGTGACTTCAACCCCTA 486
Db 579 AGGAGGAGACCCAGGAGCAGAGTGGGAGAGGTTGGCCAGCTGTGACTTCAACCCCTA 638
QY 487 AGGCGAGCAAGCAATGTAAGACGCTGCCCGCTGCGCTCGGTGCTCAT 535
Db 639 AGAGCAGCAAGCAATGTAAGACGCTGCCCGCTGCGCTCGGTGCTCAT 687
RESULT 6
ADW21848
ID ADW21848 standard; cDNA; 982 BP.
XX AC ADW21848;
XX DT 10-MAR-2005 (first entry)
XX DE Rat hepatotoxicity marker gene, SEQ:227.
XX KW Toxicology screening; drug screening; gene expression;
KW expression profile; hepatotoxicity; drug-induced; hepatitis;
KW liver disease; gastrointestinal disease; gene; ss.
XX OS Rattus norvegicus.
XX PN EP1344834-A2.
XX PD 17-SEP-2003.
XX PF 04-MAR-2003; 2003EP-00004810.
XX PR 14-MAR-2002; 2002EP-00005336.
XX PR 17-JUL-2002; 2002EP-00015657.
XX PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX PI Boess F, Suter-Dick L, Wolf D;
XX WPI: 2003-723475/69.
XX GENBANK; M15883.
XX Predicting toxicity of compounds, useful in development of safe drugs, by
XX measuring the differential expression of specific genes in cells exposed
XX to test compounds.
XX Claim 2; SEQ ID NO 227; 895pp; English.
XX The invention relates to methods of predicting at least one toxic effect
XX (or toxicity progression or the mechanism of toxicity) of a compound. The
XX methods involve detecting the level of expression of at least one of a
XX set of 680 genes ADW21622-ADW22301 or at least one of a set of 17 genes

CC (including ADW22362, ADW22414 and ADW22481-ADW22483) in a tissue or cell
CC exposed to the compound, and determining whether the gene is
CC differentially expressed compared with a control tissue or cell.
CC Differential expression of the gene in the presence of the compound is
CC indicative of a toxic effect, of toxicity progression or of a specific
CC mechanism of toxicity. The toxic effect is especially hepatotoxicity,
CC particularly hepatitis, liver necrosis, protein adduct formation or fatty
CC liver. The invention also relates to sets of primers and probes specific
CC for at least two genes selected from ADW21622-ADW22301; solid supports
CC (e.g., DNA chips) and kits containing the probes; and a database
CC containing DNA sequence information and expression information for at
CC least two of the 680 genes from hepatotoxin-exposed tissues. The
CC invention is based on the determination of global changes in gene
CC expression in tissues or cells exposed to known toxins, particularly
CC hepatotoxins, and the identification of individual genes (toxicity
CC markers) that are differentially expressed on toxin exposure. The changes
CC in gene expression can be characteristic of different mechanisms of
CC hepatotoxicity mediated by various classes of compounds. Such compounds
CC include: direct acting compounds which cause damage to macromolecules,
CC especially proteins and lipids by directly interacting with them;
CC static compounds which cause an accumulation of fat in the liver; and
CC cholestatic compounds which impair bile flow or bile acid transport,
CC resulting in jaundice. The methods of the invention are useful in
CC toxicology screening for predicting the toxic effects (especially
CC hepatotoxic effects) of compounds for the development of safer drugs.
CC Sequences ADW21622-ADW22301 represent specifically claimed hepatotoxicity
CC marker genes of rat origin whose expression is altered on exposure to
CC hepatotoxins.

XX SQ Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;

Query Match 88.0%; Score 471; DB 11; Length 982;
Best Local Similarity 90.8%; Pred. No. 1.5e-115;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 CGGCCGCTTCTCGCCGAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 60
DB 99 CGGCCGCTTCTCGCCGAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 158

QY 61 TCGGGGACCTGCGCCGAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 120
DB 159 TCGGGGACCTGCGCCGAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 218

QY 121 GTTCGGAGGACATGGGAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 180
DB 219 GTTCGGAGGACATGGGAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 278

QY 181 CCGATGCTAGCTGCGGAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 240
DB 279 CCGATGCTAGCTGCGGAGGAGCGGAGATTGCTGGCATCGAATGACTCGGGTT 338

QY 241 GCAAGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
DB 339 GCAAGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 398

QY 301 CCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
DB 399 CCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 458

QY 361 GTGAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 394
DB 459 GTGAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 518

QY 395 -----GGCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 426
DB 519 AGCCAGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 578

QY 427 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 486
DB 579 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 638

QY 487 AGAGCAGCAAGCAATGTAAGACGCTGTCGCCCTCGCTCGGTCGCTCAT 535

Db 639 AGAGCAGCAAGCAATGTAAGACGCTGTCGCCCTCGCTCGGTCGCTCAT 687

RESULT 7
ADP72639

ID ADP72639 standard; DNA; 982 BP.

XX AC ADP72639;

XX 26-AUG-2004 (first entry)

DE Renal toxin progression gene marker #1228.

XX ds; toxic effect; gene expression profile; kidney tissue;
KW differential gene expression; toxicity progression; toxicity marker;
KW drug screening; toxicity assay; kidney pathology; nephritis;
KW kidney necrosis; glomerular injury; tubular injury;
KW focal segmental glomerulosclerosis.

OS Rattus norvegicus.

XX WO2004048598-A2.

XX 10-JUN-2004.

XX 24-NOV-2003; 2003WO-US037556.

XX 22-NOV-2002; 2002US-00301856.

XX (GENE-) GENE LOGIC INC.

XX Mendrick DL, Porter MW, Johnson KR, Castle A, Higgs B;
PI Elashoff M;

XX WPI; 2004-460771/43.

XX Predicting (the progression of) a toxic effect of a compound, for
XX monitoring the progression of renal disease states, comprises preparing a
XX gene expression profile of a kidney tissue or cell sample exposed to the
XX compound.

XX Claim 11; SEQ ID NO 1228; 266pp; English.

XX The invention relates to a method of predicting (the progression of) a
XX toxic effect of a compound by preparing a gene expression profile of a
XX kidney tissue or cell sample exposed to the compound and comparing the
XX gene expression profile to a database, or detecting the level of gene(s)
XX expression in a tissue or cell sample exposed to the compound, where
XX differential gene expression compared to a control indicates a toxic
XX effect (toxicity progression). The method is useful for predicting (the
XX progression of) at least one toxic effect of a compound. The genes are
XX useful as toxicity markers in drug screening and toxicity assays. The
XX methods are useful for predicting the likelihood that a compound or test
XX agent will induce various specific kidney pathologies, such as nephritis,
XX kidney necrosis, glomerular and tubular injury, or focal segmental
XX glomerulosclerosis. The methods are useful for determining the similarity
XX of a toxic response to one or more individual compounds and for
XX predicting or elucidating the potential cellular pathways influenced,
XX induced or modulated by the compound or test agent. The kit is useful for
XX predicting or modelling the toxic response of a test compound, for
XX monitoring the progression of renal disease states, for identifying genes
XX that show promise as new drug targets and for screening known and newly
XX designed drugs. This sequence corresponds to a gene marker used in the
XX method of the invention. (Note: The sequence data for this patent did not
XX form part of the printed specification, but was obtained in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences).

XX Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;

Query Match 88.0%; Score 471; DB 12; Length 982;

Best Local Similarity 90.8%; Pred. No. 1.5e-115;

Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;


```
PN US2004018527-A1.
XX
XX
XX 29-JAN-2004.
XX
XX 16-MAY-2003; 2003US-00439703.
XX
XX 17-MAY-2002; 2002US-0381141P.
XX
XX (CHAN/) CHANG J C.
XX (OCN/) O'CONNELL P.
XX
XX Chang JC, O'Connell P;
XX
XX WPI; 2004-224389/21.
XX
XX Array useful for screening patient for resistance to docetaxel comprises
XX complementary nucleic acid probes attached to solid surface.
XX
XX Claim 21; SEQ ID NO 14; 169pp; English.
XX
XX The invention relates to an array for screening a patient for resistance
XX to docetaxel comprising complementary nucleic acid probes attached to a
XX solid surface for at least 10 of the nucleic acids chosen from 91 fully
XX defined sequences as given in the specification. The array is useful for
XX screening a patient for resistance to docetaxel. The array is also useful
XX for monitoring a cancer patient receiving docetaxel therapy. The present
XX sequence represents a human cDNA differentially expressed in response to
XX docetaxel.
XX
XX Sequence 1051 BP; 209 A; 315 C; 361 G; 166 T; 0 U; 0 Other;
SQ
Query Match 84.4%; Score 451.8; DB 12; Length 1051;
Best Local Similarity 90.3%; Pred. No. 2.1e-110;
Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 1 CGCGCGCTTCTTCTGCGCCAGCAGGAGCGAGATTGCTGGCATCCAGATGACTCGGGTT 60
DB 110 CGCGCGCTTCTTCTGCGCCAGCAGGAGCGAGATTGCGAGCATAGAACGAGGGCT 269
QY 61 TCGGGGACCTCTCGCGCCAGCAGGAGCGAGATTGCTGGCATCCAGATGACTCGGGTT 120
DB 270 TCGGGGACCTCTCGCGCCAGCAGGAGCGAGATTGCGAGCATAGAACGAGGGCT 329
QY 121 GTTTCGAGACATGGGGACCTACAGTCAATGGAGATGTGTTTCAGGAGCTTACGGGCTG 180
DB 330 GTTTCGAGACATGGGGACCTACAGTCAATGGAGATGTGTTTCAGGAGCTTACGGGCTG 389
QY 181 CCGATGGCTACGCTCGATTGCGCGGCGGACAGGTTGACTCAGGAGCTTGAGAGCATCC 240
DB 390 CTGATGGCTACGCGAGCATTTGCCAGGCTGACAGGCTGACCCAGGAGCTTGAGAGCATCC 449
QY 241 GCAAGTGGAGAGGAGCAGAGAGAGAAAGGCTGCGAGGTTGATGCTGCTCGAAGGTA 300
DB 450 GCAAGTGGGAGGAGGAGGAGGAAACGGCTGCAAGAGCTGGATGCTGATCTAAGGTCA 509
QY 301 CGGAACAGGAGTGGCGGAGAGAGGCCAAAGAAAGCTTGGAGAGTGGAAACCGGCGCAA 360
DB 510 CGGAACAGGAAATGGCGGAGAGAGGCCCAAGAGAGCTTGGAGAGTGGAAACCGGCGCAA 569
QY 361 GTCAACAGTGTGAGAGAGAACAGATCAACAGGAGCTCGAAGAGGCTTTTGTGAAAG 420
DB 570 GTGAAACAGTGTGAGAGAGAACAGATCAACAGGAGCTCGAAGAGGCTTTTGTGAAAG 629
QY 421 AATCCAGGAGGAGAGAGCCAGGAGCAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCA 480
DB 630 AATCCAGGAGGAGAGAGCCAGGAGCAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCA 689
QY 481 ACCCTAAGAGCAGAGCAATGTAAAGCTGTTCGCGCTCGCTCGGTGCTCAT 535
DB 690 ACCCAAGAGCAGAGCAGTGTCAAGATGTGTTCGCGCTCGCTCGGTGCTCAT 744
RESULT 11
```

```
ADX07445
ID ADX07445 standard; DNA; 1051 BP.
XX
XX ADX07445;
AC
XX
XX 21-APR-2005 (first entry)
DT
XX
XX Cyclin-dependent kinase modulation biomarker DNA SEQ ID NO 2011.
DB
XX
XX Cytostatic; cyclin-dependent kinase; cdk; biomarker; gene; ds.
KW
XX
XX Homo sapiens.
OS
XX
XX WO2005012875-A2.
PN
XX
XX 10-FEB-2005.
PD
XX
XX 29-JUL-2004; 2004WO-US024424.
PF
XX
XX 29-JUL-2003; 2003US-0490890P.
PR
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
PA
XX
XX Li M, Rupnow BA, Webster KR, Jackson DG, Wong TW;
PI
XX
XX WPI; 2005-163068/17.
DR
XX
XX P-PSDB; ADX07446.
XX
XX Biomarkers useful for predicting or determining the response of a mammal
XX to a cancer treatment comprising administration of a modulator of cyclin-
XX dependent kinase activity.
XX
XX Claim 5; SEQ ID NO 2010; 141pp; English.
XX
XX This invention describes a novel method of predicting or determining
XX whether a mammal will respond or is responding to an anti-cancer agent
XX that modulates cyclin-dependent kinase (cdk) activity. The method
XX comprises measuring the level of one or more biomarkers selected from
XX 2774 biomarkers given in the specification (nucleotide sequence SEQ ID
XX NO:1246 (Genbank EST W28729) is especially preferred). The method of the
XX invention is utilized in a kit for determining or predicting whether
XX patient would be susceptible or resistant to treatment by an agent
XX modulating cdk activity. The invention also describes a method for
XX utilizing individualized genetic profiles for treating diseases and
XX disorders based on patient's response and molecular level, specialized
XX against the biomarkers and a cell culture model to identify biomarkers.
XX microarrays comprising the biomarkers described, antibodies directed
XX against the biomarkers and a cell culture model to identify biomarkers.
XX The cdk modulator is preferably N-5-[(5-(1,1-dimethylethyl)-2-
XX oxazolyl)methyl]thiol]-2-thiazolyl-4-piperidine carboxamide, 0.5-L-
XX tartaric acid salt. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. This
XX sequence encodes a biomarker used in the method of the invention.
XX
XX Sequence 1051 BP; 209 A; 315 C; 361 G; 166 T; 0 U; 0 Other;
SQ
Query Match 84.4%; Score 451.8; DB 14; Length 1051;
Best Local Similarity 90.3%; Pred. No. 2.1e-110;
Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 1 CGCGCGCTTCTTCTGCGCCAGCAGGAGCGAGATTGCTGGCATTCGAGATGACTCGGGTT 60
DB 210 CGCGCGCTTCTTCTGCGCCAGCAGGAGCGAGATTGCGAGCATAGAACGAGGGCT 269
QY 61 TCGGGGACCTCTCGCGCCAGCAGGAGCGAGATTGCTGGCATTCGAGATGACTCGGGTT 120
DB 270 TCGGGGACCTCTCGCGCCAGCAGGAGCGAGATTGCGAGCATAGAACGAGGGCT 329
QY 121 GTTTCGAGACATGGGGACCTACAGTCAATGGAGATGTGTTTCAGGAGGCTTACGGGCTG 180
DB 330 GTTTCGAGACATGGGGACCTACAGTCAATGGAGATGTGTTTCAGGAGGCTTACGGGCTG 389
QY 181 CCGATGGCTACGCTCGATTGCGCGGCGGACAGGTTGACTCAGGAGCTTGAGAGCATCC 240
```


XX WO200177290-A2.
PN
XX
XX 18-OCT-2001.
PD
XX
XX 29-MAR-2001; 2001WO-US010295.
PP
XX
XX 06-APR-2000; 2000US-0194941P.
PR
XX
XX (GEMY) GENETICS INST INC.
PA
XX
XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
XX WPI; 2002-179323/23.
DR
XX
XX Six hundred and twenty five polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for treating
PT immune deficiencies and disorders such as autoimmune disorders.
XX
XX Claim 1; Page 129; 339pp; English.
PS
XX
XX The invention relates to 625 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins, their complements and sequences that hybridise to them. Also
CC included are a vector comprising the polynucleotide, a host cell
CC transformed with the vector, the proteins encoded by the polynucleotides,
CC antibodies that bind to the proteins and identification of modulators of
CC the proteins or the expression of the polynucleotide. The polynucleotides
CC can be used as probes for the identification and isolation of full length
CC cDNA and genomic DNA. The polynucleotides and proteins can also be used
CC as nutritional supplements. The protein is useful in the treatment of
CC various immune deficiencies and disorders such as viral infections,
CC bacterial infections, fungal infections, autoimmune disorders (e.g.
CC rheumatoid arthritis, multiple sclerosis, autoimmune thyroiditis and
CC diabetes) and allergic reactions and conditions (e.g. asthma). They are
CC also useful for treating neurodegenerative diseases (e.g. Alzheimer's
CC disease, Parkinson's disease), liver fibrosis, coagulation disorders
CC (e.g. haemophilia), inflammatory disorders (e.g. Crohn's disease) and
CC tumours. They are also useful for tissue regeneration, for wound healing
CC and in the treatment of burns, incisions and ulcers. The proteins are
CC also useful for regulating haematopoiesis, for treating myeloid or
CC lymphoid cell deficiencies. The present sequence is one of the 625 cDNA
CC sequences encoding a secreted protein
XX
XX Sequence 894 BP; 154 A; 280 C; 254 G; 206 T; 0 U; 0 Other;
SQ

Query Match 71.2%; Score 380.8; DB 6; Length 894;
Best Local Similarity 81.8%; Pred. No. 1.8e-91;
Matches 476; Conservative 0; Mismatches 52; Indels 54; Gaps 1;
1

QY 8 CTTCTCTGCCCGCCAGCAGGAGCGAGATTGCTGGCATCGAGAGATGACTCGGGTTTCGGGCG 67
Db CTTCTCTGCCCGCCAGCAGGAGCGAGATTGCTGGCATCGAGAGATGACTCGGGTTTCGGGCG 835
QY 68 ACTTGGCGCCAGCAGGTTGGCTCTTGGCGAGCCGAGCTCGCGAGCGGGGGTGTTCGGA 127
Db ACTTGGCGCCAGCAGGTTGGCTCTTGGCGAGCCGAGCTCGCGAGCGGGGGTGTTCGGA 775
QY 128 GCACATGGGGACTACATGTCATGAGATGTGTTTTCAGGAGGCTTAACGGCCCTGCCGATGG 187
Db GCACATGGGGAGCACACAGTCAATGAGATGTGTTTTCAGGAGGCTTAACGGCCCTGCCGATGG 715
QY 188 CTACGCTGCGATTGCCCGCCGAGCAGAGTTGACTCAGGAGCCTGAGAGCATCCGCAAGTG 247
Db CTACGCGAGCCATTGCCCGCCGCTGACAGCTGACCCAGGAGCCTGAGAGCATCCGCAAGTG 655
QY 248 GAGAGAGGACGAGAGAAAGCGTCGAGAGTTGGATGCTGCTCGAAGGTGACCGCAACA 307
Db GCGAGAGGACGAGAGAAAGCGTGCAGAGCTGGATGCTGCTTAAGGTGACCGCAACA 595
QY 308 GAGTGGCGGGAGAGGCCAAGAAAGACCTGGAGGAGTGGAAACCGCCGCAAGTGAACA 367
|

Db 594 GGAATGCGGGAGAGAGGCCCAAGAGGACCTTGGAGAGTGGAACCGAGCCAGAGTGAACA 535
QY 368 GGTGAGAAGAAACAAGATCAACAACAG----- 394
|
Db 534 AGTAGAGAGNAACAAGATCAACAACCGGATCGCTGACAAAGACATTTCTACCAGAGCCAGA 475
QY 395 -----GGCATCGAAGAGCGCTTTTGTGAAAGATTTTGTGAAAGATCAAGGAGGA 433
|
Db 474 TGCTGATATCATCGGCTACGTGGCATCGAGAGCGCTTTTCGTGAAGGAATCCNAGGAGGA 415
QY 434 GACCCCGGACACAGAGTGGGAGAGAGTGGCCAGCTGTGTGACTTCAACCCCTAAGAGCAG 493
Db 414 GACCCCGGACACAGAGTGGGAGAGAGTGGCCAGCTGTGTGACTTCAACCCCTAAGAGCAG 355
QY 494 CAAGCAATGTAAAGACGTCCTCCCGCTCGCGCTCGGCTGCTCAT 535
|
Db 354 CAAGCAGTGCNAGATGTGTCCCGCTCGGCTCGGCTGCTCAT 313
|
RESULT 14
ABK34409/C
ID ABK34409 standard; cDNA; 907 BP.
XX
AC ABK34409;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA for novel secreted protein, SEQ ID 178.
XX
KW Human; ss; gene; secreted protein; immune deficiency; viral infection;
KW bacterial infection; fungal infection; autoimmune disorder; burn;
KW rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KW diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW lymphoid cell deficiency.
XX Homo sapiens.
XX
PN WO200177290-A2.
XX
PD 18-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US010295.
XX
XX 06-APR-2000; 2000US-0194941P.
XX (GEMY) GENETICS INST INC.
XX
XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
XX WPI; 2002-179323/23.
XX
XX Six hundred and twenty five polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for treating
PT immune deficiencies and disorders such as autoimmune disorders.
XX
XX Claim 1; Page 118-119; 339pp; English.
PS
XX
XX The invention relates to 625 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins, their complements and sequences that hybridise to them. Also
CC included are a vector comprising the polynucleotide, a host cell
CC transformed with the vector, the proteins encoded by the polynucleotides,
CC antibodies that bind to the proteins and identification of modulators of
CC the proteins or the expression of the polynucleotide. The polynucleotides
CC can be used as probes for the identification and isolation of full length
CC cDNA and genomic DNA. The polynucleotides and proteins can also be used
CC as nutritional supplements. The protein is useful in the treatment of
CC various immune deficiencies and disorders such as viral infections,
CC bacterial infections, fungal infections, autoimmune disorders (e.g.

Qy 457 AGGTGGCCCGCTGTGTGACTTCAACCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCC 516
|||||
Db 404 AGGTGGCCCGCTGTGTGACTTCAACCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCC 463
|||||

Qy 517 GCCTGCGCTCGGTGCTCAT 535
|||||
Db 464 GCCTGCGCTCGGTGCTCAT 482
|||||

Search completed: January 5, 2006, 19:41:17
Job time : 492 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 5, 2006, 19:14:00 ; Search time 3751 Seconds
(without alignments)
6673.178 Million cell updates/sec

Title: US-10-664-705-145
Perfect score: 535
Sequence: 1 cggcgctctctgcccag.....cgctgctggtgctcat 535

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	535	100.0	535	1	AA874955
2	535	100.0	682	1	AW914351
3	506.2	94.6	844	6	CA322501
4	506.2	94.6	1577	4	AK029382
5	503	94.0	651	2	BB614138
6	499.4	93.3	1026	1	AV129358
7	498.2	93.1	575	7	CK344101
8	498.2	93.1	612	7	CK344230
9	493.2	92.2	557	2	BF467127
10	490.6	91.7	718	3	BI659300
11	472.8	88.4	502	1	AI463714
12	471.8	88.2	663	5	BY753851
13	471.4	88.1	569	2	BE854705
14	467.4	87.4	496	2	BF227172
15	467.2	87.3	476	1	AI104105
16	466.2	87.1	523	2	BE853395
17	459.8	85.9	496	2	BF720742
18	451.8	84.4	564	1	AW575961
19	451.8	84.4	1071	3	BM563208
20	451	84.3	451	2	BE098738
21	447.8	83.7	487	2	BE852625
22	445.4	83.3	782	3	BP438058

23	445.4	83.3	838	3	BP440642
24	443.8	83.0	817	3	BP143717
25	442.2	82.7	796	8	CK091448
26	442.2	82.7	825	2	BQ177244
27	440.8	82.4	678	2	BG470345
28	440.6	82.4	644	5	BY739827
29	440.2	82.3	775	3	BI414098
30	439.2	82.1	659	7	CN409242
31	436.2	81.5	660	5	BY737187
32	435.4	81.4	759	3	BP438499
33	434.2	81.2	803	7	CJ030794
34	434.2	81.2	826	8	DN865071
35	434.2	81.2	835	8	DN103678
36	432.6	80.9	860	8	DN114523
37	432.6	80.9	873	8	DN120155
38	432.4	80.8	605	2	BF286069
39	430.2	80.4	516	2	BE646508
40	429.8	80.3	514	1	AI128183
41	429.4	80.3	675	7	CK982463
42	429.4	80.3	685	7	CK940151
43	429.4	80.3	724	7	CK957149
44	428.4	80.1	673	7	CK940272
45	428.2	80.0	1058	3	BM558124

ALIGNMENTS

RESULT 1
AA874955
LOCUS

DEFINITION

AA874955 535 bp mRNA linear EST 05-FEB-1999
UI-R-E0-ci-g-06-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
UI-R-E0-ci-g-06-0-UI 3' similar to gi|203360|gb|M19262|RATCBR3 Rat
clathrin light chain (LCB3) mRNA, complete cds, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

COMMENT

AA874955 535 bp mRNA linear EST 05-FEB-1999
UI-R-E0-ci-g-06-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
UI-R-E0-ci-g-06-0-UI 3' similar to gi|203360|gb|M19262|RATCBR3 Rat
clathrin light chain (LCB3) mRNA, complete cds, mRNA sequence.
AA874955
AA874955.1 GI:4230837
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 535)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
On Mar 20, 1998 this sequence version replaced gi:2979903.
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa, IA 52242, USA
375 Newton Road, 4156 MEBRP, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Oligo-dT track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics
Seq primer: M13 Forward.
Location/Qualifiers
1..535
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-ci-g-06-0-UI"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-E0"
/note="Vector: pTT73D-pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; This library

consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

```
ORIGIN
Query Match      100.0%; Score 535; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 4.6e-121;
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGCGGCTTCTCGCCAGCAGGAGCGAGATGCTGGCATCGAGATGACTCGGGTT 60
Db 1 CGGCGGCTTCTCGCCAGCAGGAGCGAGATGCTGGCATCGAGATGACTCGGGTT 60

Qy 61 TCGGGGACCTTCGCGCCAGCCAGGTGGCTCTGCGCAGCCGGACTCGCAGCGGGGTG 120
Db 61 TCGGGGACCTTCGCGCCAGCCAGGTGGCTCTGCGCAGCCGGACTCGCAGCGGGGTG 120

Qy 121 GTTCGGAGGACATGGGACTACAGTCAATGAGATGTTTTCAGGAGCTAAACGGGCTG 180
Db 121 GTTCGGAGGACATGGGACTACAGTCAATGAGATGTTTTCAGGAGCTAAACGGGCTG 180

Qy 181 CCGATGCTACGTCGCGATTCGCCAGCGGACAGGTTGACTCAGGAGCTGAGAGCATCC 240
Db 181 CCGATGCTACGTCGCGATTCGCCAGCGGACAGGTTGACTCAGGAGCTGAGAGCATCC 240

Qy 241 GCAAGTGGAGAGGAGCAGAGAAAGGCTGCGAGGATTCGATGCTCGAAGGTGA 300
Db 241 GCAAGTGGAGAGGAGCAGAGAAAGGCTGCGAGGATTCGATGCTCGAAGGTGA 300

Qy 301 CCGAACAGGAGTGGCGGAGAGGCCCAAAAAGACCTGGAGAGTGGAAACAGCGCCAAA 360
Db 301 CCGAACAGGAGTGGCGGAGAGGCCCAAAAAGACCTGGAGAGTGGAAACAGCGCCAAA 360

Qy 361 GTGAACAGGTTGAGAGAACAGATCAACACAGGGCATCGGAAGAGCTTTTGTGAAG 420
Db 361 GTGAACAGGTTGAGAGAACAGATCAACACAGGGCATCGGAAGAGCTTTTGTGAAG 420

Qy 421 AATCCAGGAGGAGACCCAGGACAGAGTGGGAGAGGTGGCCAGCTGTGACTTCA 480
Db 421 AATCCAGGAGGAGACCCAGGACAGAGTGGGAGAGGTGGCCAGCTGTGACTTCA 480

Qy 481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTCGCTCGGTGCTCAT 535
Db 481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTCGCTCGGTGCTCAT 535
```

```
RESULT 2
AW914351      682 bp mRNA linear EST 25-MAY-2000
LOCUS          EST345655 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
DEFINITION    RGIAD81 5' end, mRNA sequence.
ACCESSION     AW914351
VERSION       AW914351.1 GI:8080027
KEYWORDS      EST.
SOURCE        Rattus sp.
ORGANISM      Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
```

```
1 (bases 1 to 682)
Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
Rat Genome Project: Generation of a Rat EST (Rat) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel.: (301)-838-3529
Fax: (301)-838-0208
```

Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information
Seq primer: M13 Reverse.

```
FEATURES
source
1..682
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="taxon:10118"
/clone="RGIAD81"
/clone_lib="Normalized rat brain, Bento Soares"
/notes="Organ: brain; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"
```

ORIGIN

```
Query Match      100.0%; Score 535; DB 1; Length 682;
Best Local Similarity 100.0%; Pred. No. 4.8e-121;
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGCGGCTTCTCGCCAGCAGGAGCGAGATGCTGGCATCGAGATGACTCGGGTT 60
Db 141 CGGCGGCTTCTCGCCAGCAGGAGCGAGATGCTGGCATCGAGATGACTCGGGTT 200

Qy 61 TCGGGGACCTTCGCGCCAGCCAGGTGGCTCTGCGCAGCCGGACTCGCAGCGGGGTG 120
Db 201 TCGGGGACCTTCGCGCCAGCCAGGTGGCTCTGCGCAGCCGGACTCGCAGCGGGGTG 260

Qy 121 GTTCGGAGGACATGGGACTACAGTCAATGAGATGTTTTCAGGAGCTAAACGGGCTG 180
Db 261 GTTCGGAGGACATGGGACTACAGTCAATGAGATGTTTTCAGGAGCTAAACGGGCTG 320

Qy 181 CCGATGCTACGTCGCGATTCGCCAGCGGACAGGTTGACTCAGGAGCTGAGAGCATCC 240
Db 321 CCGATGCTACGTCGCGATTCGCCAGCGGACAGGTTGACTCAGGAGCTGAGAGCATCC 380

Qy 241 GCAAGTGGAGAGGAGCAGAGAAAGGCTGCGAGGATTCGATGCTCGAAGGTGA 300
Db 381 GCAAGTGGAGAGGAGCAGAGAAAGGCTGCGAGGATTCGATGCTCGAAGGTGA 440

Qy 301 CCGAACAGGAGTGGCGGAGAGGCCCAAAAAGACCTGGAGAGTGGAAACAGCGCCAAA 360
Db 441 CCGAACAGGAGTGGCGGAGAGGCCCAAAAAGACCTGGAGAGTGGAAACAGCGCCAAA 500

Qy 361 GTGAACAGGTTGAGAGAACAGATCAACACAGGGCATCGGAAGAGCTTTTGTGAAG 420
Db 501 GTGAACAGGTTGAGAGAACAGATCAACACAGGGCATCGGAAGAGCTTTTGTGAAG 560

Qy 421 AATCCAGGAGGAGACCCAGGACAGAGTGGGAGAGGTGGCCAGCTGTGACTTCA 480
Db 561 AATCCAGGAGGAGACCCAGGACAGAGTGGGAGAGGTGGCCAGCTGTGACTTCA 620

Qy 481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTCGCTCGGTGCTCAT 535
Db 621 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTCGCTCGGTGCTCAT 675
```

RESULT 3

```
CA322501      844 bp mRNA linear EST 09-JUL-2003
LOCUS          UI-M-FX0-cch-k-13-0-UI.r1 NIH BMAP_FX0 Mus musculus cDNA clone
DEFINITION    IMAGE:6819686 5', mRNA sequence.
ACCESSION     CA322501
VERSION       CA322501.1 GI:24540599
KEYWORDS      EST.
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 844)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
```

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

FEATURES

Location/Qualifiers

1. .844

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:6819686"

/tissue_type="whole brain"

/dev_stage="embryo 13.5,14.5,16.5,17.5dpcc"

/lab_host="DHI08 (T1 phage resistant)"

/clone_lib="NIH_BMAP_FX0"

/note="Organ: Brain; Vector: pYX-Asc; Site_1: Ecor I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Query Match	94.6%	Score 506.2	DB 6	Length 844
Best Local Similarity	96.6%	Pred. No. 6.2e-114	Indels 0	Gaps 0
Matches 517	Conservative 0	Mismatches 18	Indels 0	Gaps 0
Qy	1	CGGCGGCTTCTCTGGCCAGCAGAGAGCGAGATGTGTGGCATCGAGAATGACTCGGGTT	60	
Db	78	CGGCGGCTTCTCTGGCCAGCAGGAAAGCGAGATTGTCTGGCATCGAGATGACCCGGGTT	137	
Qy	61	TCGGGGGACCTGCGCCGACAGGTGGCTCTGGCGACCCGGGACTCGCGAGCGGGGTG	120	
Db	138	TCGGGGGACCTGCGCCGACAGGTGGGCTCTGGCGACCCGGGACTCGCGAGCGGGGTG	197	
Qy	121	GTTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGAGGCTAACGGGCTG	180	
Db	198	GTTTCAGAGGACATGAGTACTACAGTCAATGGAGATGTGTTTCAGAGGCTAACGGGCTG	257	
Qy	181	CCGATGGGTACGCTGGCATGTCCGACGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC	240	
Db	258	CCGATGGGTACGCTGGCATGTCCGACGGCGGACAGGTTGACTCAGGAGCCGAGAGCATCC	317	
Qy	241	GCAAGTCGAGAGAGGAGCAGAGAAAGGCTGCAGGAGTTGATGCTGCTCCCTCGAAGGTGA	300	
Db	318	GGAAGTCGAGAGAGGAGCAGAGAAAGGCTGCAGGAGTTAGACGCTGCTCCGAGGTGA	377	
Qy	301	CCGAAACAGGAGTGGCGGAGAGGCGCAAAAAGACCTGGAGGAGTGGAAACGAGCGCCAAA	360	
Db	378	CCGAAACAGGAGTGGCGGAGAGGCGCAAAAAGACCTGGAGGAGTGGAAACGAGCGCCAAA	437	
Qy	361	GTGAACAGGTTGGAAGAACAGATCAACAACAGGGCATCGGAAGAGGCTTTGTGAAG	420	
Db	438	GTGAACAGGTTGGAAGAACAGATCAACAACAGGGCATCGGAAGAGGCTTTGTGAAG	497	
Qy	421	AATCCNAGGAGGAGACCCGAGGCAACAGATGGGAAAGGTGGCCGACGCTGTGTGACTTCA	480	

Db 498 AATCCAGAGGAGGACCCCGGCGCACAGAGTGGGAGAGGTAGCCCGAGCTGTGTGACTTCA 557

Qy 481 ACCCTTAAGAGCAGCAAGCAATGTTAAAGACGTGTCCCGCTCGCTCGGTGCTCAT 535

Db 558 ATCCAGAGAGGAGCAGCAGTGTAAAGATGTGTCCCGCTCGCTCGGTGCTCAT 612

RESULT 4

AK029382

LOCUS

DEFINITION

Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched library, clone:483419B16 product:clathrin, light polypeptide (lcb), full insert sequence.

AK029382

VERSION

AK029382.1 GI:26325351

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636

2 Carninci, P. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159

JOURNAL

REFERENCE

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsuura, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861

JOURNAL

REFERENCE

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1577)

JOURNAL

REFERENCE

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saigo, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sugabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gscc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>

URL: <http://fantom.gsc.riken.jp/>

FEATURES
source

```
1..1577
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="C57BL/6J"
    /db_xref="PANTOM.DB:4833419E16"
    /db_xref="taxon:10090"
    /clone="4833419E16"
    /tissue_type="head"
    /clone_lib="RIKEN full-length enriched mouse cDNA library"
    /dev_stage="0 day neonate"
    /dev_start=1
    /protein_id="BAC26430.1"
    /db_xref="GI:2632532"
    /translation="MASEDFGFSSESGAPEAEDPAAPLAQOESEIAGIENDPFG
    GAPAAQVASQPLGASGAGSEDSMTTVNGDFVQANGPADYAAIAQAORLTQPEPS
    IIRWREKQLQLDAASKVTEQREKAKDLLEWNRQSEQRKNNRASEA
    FVKSKEETPGTEWKEVAQLCDFNPKSKQCKDVSRLSVLSMLKQPLSR"
    /polyA_signal
    1553..1558
    /note="putative"
    /polyA_site
    1577
    /note="putative"
```

ORIGIN

```
Query Match          94.6%; Score 506.2; DB 4; Length 1577;
Best Local Similarity 96.6%; Pred. No. 6.9e-114;
Matches 517; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1 CGGCCGCTTCTCGCCAGCAGGAGCGGAGATTGCTGGCATCGAGATGACTCGGGTT 60
Db 80 CGGCCGCTTCTCGCCAGCAGGAGCGGAGATTGCTGGCATCGAGATGACTCGGGTT 139

Qy 61 TCGGGGACCTGCGCCAGCCAGCGTGGCTCTGCGCAGCCCGGACTCGGAGCGGGGTG 120
Db 140 TCGGGGACCTGCGCCAGCCAGCGTGGCTCTGCGCAGCCCGGACTCGGAGCGGGGTG 199

Qy 121 GTTCGGAGCATCGGGACTACGTCAATGAGATGCTTTCAGGAGCTTAACGGGCTG 180
Db 200 GTTCAGAGGACATGAGTACTACGTCAATGAGATGCTTTCAGGAGCTTAACGGGCTG 259

Qy 181 CCGATGGCTAGCTGCGATTGCCAGGCGGAGCAGGTTCGACTCAGGAGCTTCAGAGCATCC 240
Db 260 CCGATGGCTAGCTGCGATTGCCAGGCGGAGCAGGTTCGACTCAGGAGCTTCAGAGCATCC 319

Qy 241 GCAAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 320 GCAAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 379

Qy 301 CCGACAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 380 CCGACAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 439

Qy 361 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Db 440 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 499
```

```
Qy 421 AATCCAAGGAGGAGACCCAGGACAGAGTGGGAGAAAGGTGGCCAGCTGTGTCACTTCA 480
Db 500 AATCCAAGGAGGAGACCCAGGACAGAGTGGGAGAAAGGTAGCCAGCTGTGTCACTTCA 559
Qy 481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTGCGTCCGGTGTCTAT 535
Db 560 ATCCAAGAGCAGCAAGCAGTGTAAAGATGTGTCCCGCTGCGTCCGGTGTCTAT 614
```

RESULT 5

BB614138

LOCUS

DEFINITION

musculus cDNA clone 4833419E16 5', mRNA sequence.

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 651)

Atakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,

Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,

Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,

Onno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,

Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,

Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gscc.riken.jp, URL: <http://genome.gsc.riken.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,

Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.

and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,

Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I.,

Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and

Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences. Mamm. Genome 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.jp/>) for

further details.

e mouse tissues.

FEATURES

Location/Qualifiers

1..651

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="4833419E16"


```

Qy 63 GGGGCACTGCGCCAGCCAGGTGGCTCTGGCAGCCCGGACTCGCGAGCGGGGTGGT 122
Db 238 GGGGCACTGCGCCAGCCAGGTGGCTCTGGCAGCCCGGACTCGCGAGCGGGGTGGT 297
Qy 123 TCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGCGCTGCC 182
Db 298 TCAGAGGACATGAGTACTACAGTCATGAGATGTGTTTCAGGAGGCTTAACGGCGCTGCC 357
Qy 183 GATGCTACGTGCGATTGCCAGGCGACAGGTGATCTAGGAGGCTTGAGAGATCGGC 242
Db 358 GATGCTACGTGCGATTGCCAGGCGACAGGTGATCTAGGAGGCTTGAGAGATCGGC 417
Qy 243 AAGTGGAGAGGAGCAGAGAAAGGCTGAGGAGTTGGATGCTGCTCGAAGGTGACC 302
Db 418 AAGTGGAGAGGAGCAGAGAAAGGCTGAGGAGTTAGACGCTGCTCGAAGGTGACC 477
Qy 303 GAACAGGAGTGGCGGAGAGCCCAAAAGACCTGGAGGAGTGGAAACAGCGCCAAAGT 362
Db 478 GAACAGGAGTGGCGGAGAGCCCAAGAAAGCTGGAGGAGTGGAAACAGCGCCAAAGT 537
Qy 363 GAACAGGTTGAGAGAAACAAGATCAACACAGGCGATCGGAGAGGCTTTTGTGAAGAA 422
Db 538 GAACAGGTTGAGAGAAACAAGATCAACACAGGCGATCGGAGAGGCTTTTGTGAAGAA 597
Qy 423 TCCAGGAGGAGACCCAGGCGACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAAC 482
Db 598 TCCAGGAGGAGACCCAGGCGACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAAC 657
Qy 483 CCTAAGAGCAGCAGCAATGTAAAGACGTGTCCCGCTCGCTCGGTGCTCAT 535
Db 658 CCCAGAGCAGCAGCAGTGTAAAGATGTGTCCCGCTCGCTCGGTGCTCAT 710

```

```

RESULT 7
CK344101
LOCUS
DEFINITION K0835B11-3 NTA Mouse 8.5-dpc Whole Embryo cDNA Library (Long) Mus
musculus cDNA clone NIA:K0835B11 IMAGE:30081718 3', mRNA sequence.
ACCESSION CK344101
VERSION CK344101.1 GI:40299714
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 575)
Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
1154199
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: K0835 row: B column: 11
Seg primer: -21M13 Forward
High quality sequence stop: 575
POLYA=Yes.
Location/Qualifiers
1..575
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6J"
/db_xref="niaEST:K0835B11-3"
/db_xref="taxon:10090"
/clone="NIA:K0835B11 IMAGE:30081718"
/tissue_type="whole embryo including extraembryonic
tissues at 8.5-days postcoitum"

```

```

/dev stage="8.5-days postcoitum"
/lab_host="DH10B"
/clone_lib="NIA Mouse 8.5-dpc Whole Embryo cDNA Library
(Long)"
note="vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001)). [PMID: 11544199]. Total RNAs were
extracted from a pool of 13 embryos at 8.5-days
postcoitum. Double-stranded cDNAs were synthesized with an
Oligo(dT) primer [Invitrogen]:
5'-PGACTAGTCTAGATCGGAGCGCGCCCTTTTCTTTTCTTTT-3' from
9.1 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Tag polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.5 kb. The library was constructed
by Yulan Piao (NIA)."

```

ORIGIN

```

Query Match      93.1%; Score 498.2; DB 7; Length 575;
Best Local Similarity 96.6%; Pred. No. 5.4e-112;
Matches 509; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 9 TTCCTGGCCAGCAGGAGCGGAGATTGCTGGCATCGAAGATGACTCGGGTTTCGGGGCA 68
Db 1 TTCCTGGCCAGCAGGAGCGGAGATTGCTGGCATCGAAGATGACTCGGGTTTCGGGGCA 60
Qy 69 CCTCGCCCGCAGCAGGTTGGCTCTGCGCAGCCCGGACTCGCGAGCGGGGTGGTTCGGAG 128
Db 61 CCTCGCCCGCAGCAGGTTGGCTCTGCGCAGCCCGGACTCGCGAGCGGGGTGGTTCAGAG 120
Qy 129 GACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACCGGCTTCGCGATGCG 188
Db 121 GACATGAGTACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACCGGCTTCGCGATGCG 180
Qy 189 TACGCTGGATTGCCAGGCGGACAGTGTGACTCAGGAGCTTGAGAGCATCGGCAAGTGG 248
Db 181 TACGCTGGATTGCCAGGCGGACAGTGTGACTCAGGAGCTTGAGAGCATCGGCAAGTGG 240
Qy 249 AGAGAGGAGCAGAGAGAAAGGCTCGCAGGATGTGCTCGAAGGTGACCGGAACAG 308
Db 241 AGAGAGGAGCAGAGAGAAAGGCTCGCAGGATGTGAGAGTTCAGAGTTCGACCGAACAG 300
Qy 309 GAGTGGCGGAGAGGCGCAAAAAAGACCTGGAGAGTGGAAACAGCGCCAAAGTGAACAG 368
Db 301 GAGTGGCGGAGAGGCGCAAAAAAGACCTGGAGAGTGGAAACAGCGCCAAAGTGAACAG 360
Qy 369 GTTGAGAGAGAACAGATCAACACAGGCGATCGGAGAGGCTTTTGTGAAGATCCCAAG 428
Db 361 GTTGAGAGAGAACAGATCAACACAGGCGATCGGAGAGGCTTTTGTGAAGATCCCAAG 420
Qy 429 GAGGAGAGCCCGCAGGCGACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAACCCCTAAG 488
Db 421 GAGGAGAGCCCGCAGGCGACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAATCCCAAG 480
Qy 489 AGCAGCAGCAATGTAAAGACGTGTCCCGCTCGGCTCGGTGCTCAT 535
Db 481 AGCAGCAGCAATGTAAAGACGTGTCCCGCTCGGCTCGGTGCTCAT 527

```

```

RESULT 8
CK344230
LOCUS CK344230 612 bp mRNA linear EST 22-DEC-2003

```


DEFINITION K0855H11-3 NIA Mouse 8.5-dpc Whole Embryo cDNA Library (Long) Mus musculus cDNA clone NIA:K0855H11 IMAGE:30083710 3', mRNA sequence.

CK344230

ACCESSION CK344230.1 GI:40299843

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 612)

Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.

Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method

Genome Res. 11 (9), 1553-1558 (2001)

11544199

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@gsun.grc.nia.nih.gov

Plate: K0855 row: H column: 11

Seq primer: -21M13 Forward

High quality sequence stop: 612

POLYA=Yes.

Location/Qualifiers

1..612

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/db_xref="taxon:K0855H11-3"

/clone_lib="NIA Mouse 8.5-dpc Whole Embryo cDNA Library (Long)"

/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 13 embryos at 8.5-days postcoitum. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen].

5'-pGACTAGTCTAGATCGAGCGCGCCCTTTTCTTTT-3' from 9.1 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 Kb. The library was constructed by Yulan Piao (NIA)."

FEATURES

source

RESULT 9

BF467127

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

TITLE

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

889548

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEST@mail.nih.gov

Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward

POLYA=No.

Location/Qualifiers

1..557

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

ORIGIN

Query Match 93.1%; Score 498.2; DB 7; Length 612;

Best Local Similarity 96.8%; Pred. No. 5.5e-112;

Matches 509; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 9 TTCTCGCCCGCAGCAGGACGAGATGCTGGCATCGAGATGACTCGGGTTTCGGGGCA 68

Db 1 TTCTCGCCCGCAGCAGGACGAGATGCTGGCATCGAGATGACTCGGGTTTCGGGGCA 60

QY 69 CCTCGCCGACCCAGGTGGCTCTCGCAGCCCGGACTCGCAGCGGGGTGGTTCCGAG 128

Db 61 CCTCGCCGACCCAGGTGGCTCTCGCAGCCCGGACTCGCAGCGGGGTGGTTCCGAG 120

QY 129 GACATGGGGACTACAGTCAATGGAGATGTGTTTTCAGGAGGCTTAACGGGCTGCCGATGCG 188

Db 121 GACATGAGTACTACAGTCAATGGAGATGTGTTTTCAGGAGGCTTAACGGGCTGCCGATGCG 180

QY 189 TACGCTGCGATTGCGCCAGCGGACAGGTTTACTCAGGAGCCTGAGAGCATCCGCAAGTGG 248

Db 181 TACGCTGCGATTGCGCCAGCGGACAGGTTTACTCAGGAGCCTGAGAGCATCCGCAAGTGG 240

QY 249 AGAGAGGAGCAGAGAGAAAGGCTCGAGAGTTGGATGTGCTCTGAAAGTGAACCAACAG 308

Db 241 AGAGAGGAGCAGAGAGAAAGGCTCGAGAGTTAGACGCTGCTCGAAGGTGACCCGACAG 300

QY 309 GAGTGGCGGAGAGAGGCCCAAAAAGACCTCGAGAGTGGAGCCAGCCGCAAGTGAACAG 368

Db 301 GAGTGGCGGAGAGAGGCCCAAAAAGACCTCGAGAGTGGAGCCAGCCGCAAGTGAACAG 360

QY 369 GTTGAGAGAGAACAGATCAACACAGGGGCATCGGAAGAGGCTTTTGTGAAGAATCCAAG 428

Db 361 GTTGAGAGAGAACAGATCAACACAGGGGCATCGGAAGAGGCTTTTGTGAAGAATCCAAG 420

QY 429 GAGGAGACCCCGAGCAGACAGAGTGGGAGAGGTGGCCAGCTGTGTGACTTCAACCCCTAAG 488

Db 421 GAGGAGACCCCGAGCAGACAGAGTGGGAGAGGTAGCCAGCTGTGTGACTTCAACCCCTAAG 480

QY 489 AGCAGCAAGCAATGTAAAGAGCTGTCGCCCTGCGCTCGGTGCTCAT 535

Db 481 AGCAGCAAGCAATGTAAAGAGTGTGTCCCGCTGCGCTCGGTGCTCAT 527

FEATURES

source

/db_xref="taxon:10090"
 /clone="UI-M-CG0P-bre-h-11-0-UI"
 /lab host="DH10B (Life Technologies)"
 /clone lib="NIH_BMAP Ret4 S2"
 /note="Vector: p77T3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not 1; Site 2: Eco RI; The
 NIH_BMAP Ret4_S2 library is a subtracted library,
 ultimately derived from mouse retina tissue libraries at
 various stages of development. For a detailed description
 of the library from which this clone was derived, please
 visit our web site at brainest.eng.uiowa.edu. The tissue
 for this library was contributed by Dr. Xin-Yuan Fu, Yale
 University School of Medicine
 TAG_SEQ=None found"

ORIGIN
 Query Match 92.2%; Score 493.2; DB 2; Length 557;
 Best Local Similarity 96.3%; Pred. No. 9.2e-111;
 Matches 515; Conservative 0; Mismatches 19; Indels 1; Gaps 1;
 Qy 1 CGGCGGCTTCTGCGCCAGCAGGAGCGAGATTGCTGGCATCGAATGACTCGGGTT 60
 Db 2 CGGCGGCTTCTGCGCCAGCAGGAGCGAGATTGCTGGCATCGAATGACTCGGGTT 61
 Qy 61 TCGGGGACCTGCGCCAGCAGGAGCGAGATTGCTGGCATCGAATGACTCGGGTT 120
 Db 62 TCGGGGACCTGCGCCAGCAGGAGCGAGATTGCTGGCATCGAATGACTCGGGTT 121
 Qy 121 GTTCGGAGGACATGGGGACTACAGTCAATGAGATGTTTTCAGGAGCTAACGGGCTG 180
 Db 122 GTTCAGAGCATAGTACTACAGTCAATGAGATGTTTTCAGGAGCTAACGGGCTG 181
 Qy 181 CCGATGCTACGCTGCGATTGCCAGCGGACAGGTTGACTCAGAGCTTCAGAGCATCC 240
 Db 182 CCGATGCTACGCTGCGATTGCCAGCGGACAGGTTGACTCAGAGCTTCAGAGCATCC 241
 Qy 241 GCAAGTGAG 300
 Db 242 GGAAGTGAG 301
 Qy 301 CCGAACAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 Db 302 CCGAACAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
 Qy 361 GTGAACAGGTTGAG 420
 Db 362 GTGAACAGGTTGAG 421
 Qy 421 AATCAAG 480
 Db 422 AATCAAG 480
 Qy 481 ACCCTAAGCAGCAGCAAGCAATGAAGAGCTGTCGCCGCTCGCTCGGTGCTCAT 535
 Db 481 ATCCCAAGCAGCAGCAAGCAAGTGAAGATGTGTCNCGCTCGCTCGGTGCTCAT 535

RESULT 10
 BI659300/c
 LOCUS 603301689F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:534748 5',
 DEFINITION mRNA sequence.
 ACCESSION BI659300
 VERSION BI659300.1 GI:15573536
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 718)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
 COMMENT

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
 Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM11883 row: a column: 23
 High quality sequence stop: 718.
 Location/Qualifiers
 1. 718
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NMRI"
 /db_xref="taxon:10090"
 /clone="IMAGE:5347486"
 /tissue_type="tumor, gross tissue"
 /dev_stage="5 months"
 /lab host="DH10B"
 /clone lib="NCI CGAP Mam4"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Priscilla Furth,
 NIH Reference for transgenic model: Li et al., Cell Growth
 and Differentiation 7, 3-11 (1996)."

ORIGIN

Query Match 91.7%; Score 490.6; DB 3; Length 718;
 Best Local Similarity 96.4%; Pred. No. 4.2e-110;
 Matches 502; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
 Qy 15 GCCCAGCAGGAGAGCGAGATTGCTGGCATCGAATGACTCGGGTTTCGGGGCACCTGCC 74
 Db 714 GCCCAGCAGGAGAGCGAGATTGCTGGCATCGAATGACTCGGGTTTCGGGGCACCTGCC 655
 Qy 75 GCCCAGCAGGTTGGCTCTGCGCAGCCCGGACTCGCAGCGGGGTGTTTCGAGAGACATG 134
 Db 654 GCCCAGCAGGTTGGCTCTGCGCAGCCCGGACTCGCAGCGGGGTGTTTCGAGAGACATG 595
 Qy 135 GGGACTACAGTCAATGAGATGTTTTCAGAGAGCTTACCGGCTTCGCGATGGCTACGCT 194
 Db 594 AGTACTACAGTCAATGAGATGTTTTCAGAGAGCTTACCGGCTTCGCGATGGCTACGCT 535
 Qy 195 GCGATTGCCAGGCGGACAGGTTGACTCAGGAGCTGAGAGCATCCGCAAGTGGAGAGAG 254
 Db 534 GCGATTGCCAGGCGGACAGGTTGACTCAGGAGCTGAGAGCATCCGCAAGTGGAGAGAG 475
 Qy 255 GAGCAGAGAGAGAGGCTGCGAGGTTGGATGCTGCTCGAAGGTGACCCGAAACAGAGTGG 314
 Db 474 GAGCAGAGAGAGAGGCTGCGAGGTTAGACGCTGCTCGAAGGTGACCCGAAACAGAGTGG 415
 Qy 315 CGGAGAGAGGCGCAAAAGACCTGAGAGTGGAAACAGCGCCCAAGTGAACAGGTTGAG 374
 Db 414 CGGAGAGAGGCGCAAAAGACCTGAGAGTGGAAACAGCGCCCAAGTGAACAGGTTGAG 355
 Qy 375 AAGAACAGAGTCAACACAGGCGCATCGAAGAGGCTTTTGTGAAAGAAATCCAAAGAGGAG 434
 Db 354 AAGAACAGAGTCAACACAGGCGCATCGAAGAGGCTTTTGTGAAAGAAATCCAAAGAGGAG 295
 Qy 435 ACCCCAGGACAGAGTGGGAGAGGTGCCAGGCTGTGTGACTTCAACCTTAAAGAGCAGC 494
 Db 294 ACCCCAGGACAGAGTGGGAGAGGTAGTACCAGGCTGTGTGACTTCAATCCCAAGAGCAGC 235
 Qy 495 AAGCATGTAAGAGAGCTGTCCCGCTCGCTCGGTGCTCAT 535
 Db 234 AAGCAGTGTAAAGATGTGTCCCGCTCGCTCGGTGCTCAT 194

```

RESULT 11
AI463714      502 bp      mRNA      linear      EST 09-MAR-1999
LOCUS        uc45h03.x1 Soares mammary_gland_NLMG Mus musculus cDNA clone
DEFINITION   IMAGE:1400981 3' similar to gb:M20469 CLATHRIN LIGHT CHAIN B
              (HUMAN);, mRNA sequence.
ACCESSION   AI463714
VERSION     AI463714.1 GI:4317744
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 502)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            This clone is available royalty-free through LNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:912697
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            Possible reversed clone: polyT not found
            High quality sequence stop: 478.

FEATURES             source
     1..502
         /organism="Mus musculus"
         /mol_type="mRNA"
         /db_xref="taxon:10090"
         /clone="IMAGE:1400981"
         /sex="female (lactating)"
         /tissue_type="mammary gland"
         /lab_host="DH10B"
         /clone_lib="Soares mammary_gland_NLMG"
         /note="Vector: pT73D-Pac (Pharmacia) with a modified
            polylinker; 1st strand cDNA was prepared from mammary
            gland tissue from a lactating female, and was then primed
            with a Not I - oligo(dT) primer. Double-stranded cDNA was
            ligated to Eco RI adaptors (Pharmacia), digested with Not
            I and cloned into the Not I and Eco RI sites of the
            modified pT73 vector. Library is normalized. Library
            was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match      88.4%; Score 472.8; DB 1; Length 502;
Best Local Similarity 96.4%; Pred. No. 9.7e-106;
Matches 483; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 CGGCGCCCTTCTCGGCCAGCAGGAGGAGGATGCTGGCATCGAGATGACTCGGGTT 60
DB 2 CGGCGCCCTTCTCGGCCAGCAGGAGGAGGATGCTGGCATCGAGATGACTCGGGTT 61
QY 61 TCGGGGCACCTGCGCCAGCCAGGTCCTCTTGGCAGGCCCGGACTCTCGGAGCGGGGTTG 120
DB 62 TCGGGGCACCTGCGCCAGCAGGTCGCTCTGCGCAGCGGGGACTCTCGGAGCGGGGTTG 121
QY 121 GTTCGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGGCTG 180
DB 122 GTTCGAGGACATGGAGTACTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGGCTG 181
QY 181 CGGATGGCTACGCTGCGGATTCGCCAGCGGACAGGTTTGACTCAGGAGCCTGAGAGCATCC 240
DB 182 CGGATGGCTACGCTGCGGATTCGCCAGCGGACAGGTTTGACTCAGGAGCCTGAGAGCATCC 241
QY 241 GCAAGTGGAGAGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 242 GGAAGTGGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
QY 301 CCGAACAGGAGTGGCGGGAGAGAGGCCAAAAGACCTGGAGGAGTGGACCGGCCA 360

```

```

Db 302 CCGAACAGGAGTGGCGGGAGAGAGGCCAAGAACCTGGAGGAGTGGAAACCGGCCA 361
QY 361 GTGAACAGGTTGAGAACAGACAGATCAACACAGGGCATCGGAAGAGGCTTTTGTGAAG 420
DB 362 GTGAACAGGTTGAGAACAGACAGATCAACACAGGGCATCGGAAGAGGCTTTTGTGAAG 421
QY 421 AATCAAAGGAGGAGACCCCGAGGACACAGTGGGAGAGAGTGGCCCGAGCTGTGTGACTTCA 480
DB 422 AATCAAAGGAGGAGACCCCGAGGACACAGTGGGAGAGAGTGGCCCGAGCTGTGTGACTTCA 481
QY 481 ACCCTAAGAGCAGCAAGCAAT 501
DB 482 ATCCCAAGAGCAGCAAGCAGT 502

```

```

RESULT 12
BY753851
LOCUS        663 bp      mRNA      linear      EST 17-DEC-2002
DEFINITION   BV753851 RIKEN full-length enriched, osteoclast-like cell Mus
              musculus cDNA clone 142020N05 5', mRNA sequence.
ACCESSION   BV753851
VERSION     BV753851.1 GI:27185506
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 663)
AUTHORS     Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
            Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
            Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
            Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,
            Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuoka, H.,
            Batalov, S., Beisel, K. W., Blake, J. A., Bratt, D., Bruscia, V.,
            Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,
            Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
            Gariboldi, M., Giasi, C., Godzik, A., Gough, J., Grimmond, S.,
            Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
            Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,
            Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
            Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
            Numata, K., Okido, T., Pavan, W. J., Perlea, G., Pesole, G.,
            Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramchandran, S.,
            Ravasi, F., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,
            Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,
            Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
            Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
            Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
            Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
            Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
            Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
            Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
            Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
            Rogers, J., Birney, E. and Hayashizaki, Y.
            Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
            Nature 420, 563-573 (2002)
            12466851
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp URL: http://genome.gsc.riken.jp/
            Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
            Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, F., Hori, F.,
            Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
            Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,

```



```

/clone="RHEBX34"
/clone_lib="Normalized rat heart, Bento Soares"
/note="Organ: heart; Vector: pT7r3Pac; Site_1: EcoRI;
Site_2: NotI"

ORIGIN
Query Match      87.3%; Score 467.2; DB 1; Length 476;
Best Local Similarity 99.4%; Pred. No. 2.3e-104;
Matches 469; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 46 AGAATGACTCGGGTTTCGGGGCACCTGCCGCGAGCCAGGTGGCTCTCGCGAGCCCGGAC 105
Db 1 AGAATGACTCGGGTTTCGGGGCACCTGCCGCGAGCCAGGTGGCTCTCGCGAGCCCGGAC 60

Qy 106 TCGCGAGCGGGGTGGTTTCGAGGACATGGGGACTACAGTCAATCGAGATGTGTTTCAGG 165
Db 61 TCGCGAGCGGGGTGGTTTCGAGGACATGGGGACTACAGTCAATCGAGATGTGTTTCAGG 120

Qy 166 AGGCTAACGGGCTGCCGATGGCTACGCTGGATTGCCAGCGGACAGGTGACTCAGG 225
Db 121 AGGCTAACGGGCTGCCGATGGCTACGCTGGATTGCCAGCGGACAGGTGACTCAGG 180

Qy 226 AGCTGAGAGCATCCGCAAGTGGAGAGAGGAGCAGAGAAAGGCTGCAGGAGTTGGATG 285
Db 181 AGCTGAGAGCATCCGCAAGTGGAGAGAGGAGCAGAGAAAGGCTGCAGGAGTTGGATG 240

Qy 286 CTGCCTCGAAGGTGACCGAAACAGGAGTGGCGGGAGAGGCCAAAGAACCTTGAGGAGT 345
Db 241 CTGCCTCGAAGGTGACCGAAACAGGAGTGGCGGGAGAGGCCAAAGAACCTTGAGGAGT 300

Qy 346 GGAACCGCGCCAAAGTGAAACAGGTTGAGAGAGAACAGATCAACACAGGSCATCGGAAG 405
Db 301 GGAACCGCGCCAAAGTGAAACAGGTTGAGAGAGAACAGATCAACACAGGSCATCGGAAG 360

Qy 406 AGGCTTTTGTGAAGAATCCAAAGGAGAGACCCAGGACACAGAGTGGGAGAGGTGGCCC 465
Db 361 AGGCTTTTGTGAAGAATCCAAAGGAGAGACCCAGGACACAGAGTGGTAGAGAGGTGGCCC 420

Qy 466 AGCTGTGTGACTTCAACCTTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCG 517
Db 421 AGCTGTGTGACTTCAACCTTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCG 472
```

Search completed: January 5, 2006, 21:32:34
Job time : 3756 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 5, 2006, 19:27:35 ; Search time 159 Seconds

(without alignments)
5981.104 Million cell updates/sec

Title: US-10-664-705-145

Perfect score: 535

Sequence: 1 cggcgcttctggtccag.....cgctgcgtgggtcctcat 535

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:
1: /cgn2_6/prodata/1/ina/1 COMB.seq.*
2: /cgn2_6/prodata/1/ina/5 COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B COMB.seq.*
5: /cgn2_6/prodata/1/ina/H COMB.seq.*
6: /cgn2_6/prodata/1/ina/PTUS COMB.seq.*
7: /cgn2_6/prodata/1/ina/PP COMB.seq.*
8: /cgn2_6/prodata/1/ina/RE COMB.seq.*
9: /cgn2_6/prodata/1/ina/backfiles1.seq.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	78	14.6	7218	US-08-232-463-14	Sequence 14, Appl
C 2	65.2	12.2	658	US-09-385-982-327	Sequence 327, App
C 3	55.8	10.4	12695	US-09-949-016-16775	Sequence 16775, A
C 4	54.2	10.1	51259	US-08-781-891-209	Sequence 209, App
C 5	54.2	10.1	51259	US-09-618-166-209	Sequence 209, App
C 6	52.8	9.9	16442	US-08-781-891-208	Sequence 208, App
C 7	52.8	9.9	16442	US-09-618-166-208	Sequence 208, App
C 8	50.6	9.5	1926	US-09-249-585A-2	Sequence 2, Appli
C 9	50.6	9.5	1926	US-09-410-399-3	Sequence 3, Appli
C 10	50.6	9.5	2580	US-09-050-863-2	Sequence 2, Appli
C 11	50.6	9.5	2580	US-09-359-081-2	Sequence 2, Appli
C 12	50.6	9.5	5452	US-09-130-114-1	Sequence 1, Appli
C 13	50.6	9.5	8705	US-09-647-344A-14	Sequence 14, Appl
C 14	50.6	9.5	9600	US-08-910-647-1	Sequence 1, Appli
C 15	50.6	9.5	9600	US-09-620-925-1	Sequence 1, Appli
C 16	50.6	9.5	10596	US-07-884-811-15	Sequence 15, Appl
C 17	50.6	9.5	10596	US-07-885-971-15	Sequence 15, Appl
C 18	50.6	9.5	10596	US-08-087-783A-15	Sequence 15, Appl
C 19	50.6	9.5	10596	US-08-194-088B-15	Sequence 15, Appl
C 20	50.6	9.5	10596	US-08-194-087-15	Sequence 15, Appl
C 21	50.6	9.5	10596	PCT-US93-04648-15	Sequence 15, Appl
C 22	50.6	9.5	16080	US-09-724-566A-48	Sequence 48, Appl
C 23	50.6	9.5	16080	US-09-471-669A-48	Sequence 48, Appl
C 24	48.6	9.1	150394	US-09-949-016-13042	Sequence 13042, A

25	48.4	9.0	3489	2	US-08-728-323A-1	Sequence 1, Appli
26	48.4	9.0	3489	3	US-09-298-568-1	Sequence 1, Appli
27	48.4	9.0	3489	3	US-09-410-399-1	Sequence 1, Appli
28	48.4	9.0	3489	3	US-09-894-273-1	Sequence 1, Appli
C 29	48.4	9.0	32207	2	US-08-770-379-20	Sequence 20, Appl
C 30	48.4	9.0	32207	3	US-08-757-669A-20	Sequence 20, Appl
C 31	48.4	9.0	32207	3	US-09-230-371A-20	Sequence 20, Appl
C 32	47.8	8.9	669	3	US-09-669-751-79	Sequence 79, Appl
C 33	47.8	8.9	254964	3	US-09-949-016-12583	Sequence 12583, A
C 34	47.8	8.9	254964	3	US-09-949-016-17392	Sequence 17392, A
C 35	47.4	8.9	2093	3	US-10-104-047-1666	Sequence 1666, Ap
C 36	47.4	8.9	767677	3	US-09-949-016-12147	Sequence 12147, A
C 37	47.4	8.9	767677	3	US-09-949-016-17361	Sequence 17361, A
C 38	47.2	8.8	85850	3	US-09-949-016-13424	Sequence 13424, A
C 39	46.8	8.7	4092	3	US-09-566-921-115	Sequence 115, App
C 40	46.2	8.6	72549	3	US-09-949-016-16477	Sequence 16477, A
C 41	44.8	8.4	1926	3	US-09-249-585A-4	Sequence 4, Appli
C 42	44.8	8.4	1931	2	US-09-130-114-2	Sequence 2, Appli
C 43	44.6	8.3	289	3	US-09-007-005-17	Sequence 17, Appl
C 44	44.6	8.3	289	3	US-09-244-796-17	Sequence 17, Appl
C 45	44	8.2	4722	3	US-08-979-608A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, P. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 INMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 883-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-Fls

Qy	125	GGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTTCAGGAGGCTAACGGGCTCCGCA	184
Db	514	GGAGGAGGGGCGAGGACAGGAGAGGGGCGAGAGGGGCGAGGACGAGGAGGGGGCAGGA	573
Qy	185	TGGCTACGCTGCGATTGCCCTAGGCGGACAGGTTGACTTCAGGAGCCTCAGAGCATCCGCAA	244
Db	574	GGGCGAGGAGGGCGAGGACGAGAGGGGCAGGAGCAGGAGGGGCGAGGCGGCAGGA	633
Qy	245	GTGGACAGAGGACAGAGAGAAAGGCTCGAGGAGTTGGATGCTTCGTAAGGTGACCGA	304
Db	634	GGGGCGAGGACAGGAGGGGCGAGGACAGGAGAGGGGCGAGGAGGGGCGAGGAGGCGGA	693
Qy	305	ACAGGAGTGGCGGGAGAGGCCCAAAAAAGACCTCGAGGAGGTGGAAACAGGCCCAAAAGTGA	364
Db	694	GCAGGAGGGGCGAGGACGAGAGGGGCAGGAGCAGGAGGGGCGAGGACGAGGAGGGGCGAGGA	753
Qy	365	ACAGTTTGAGAGMACAGATCAACACAGGGCATCCGAGAGAGGCTTTCTGTGAAGNATC	424
Db	754	GGGCGAGGAGCAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGCA	813
Qy	425	CAAGGAGGAGACCCACAGGCACAGAGTGGGAGAGGTG	461
Db	814	GGAGGAGGGGCGAGGAGGGGCGAGGACAGGAGGAGGGG	850

RESULT 9

```

RES001.9
US-09-410-399-3
; Sequence 3, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein-Barr virus
US-09-410-399-3

```

Query Match	9.5%	Score 50.6;	DB 3;	Length 1926;
Best Local Similarity	46.9%;	Pred. No. 0.00068;		
Matches 158;	Conservative 0;	Mismatches 179;	Indels 0;	Gaps 0
QY	125	GGAGGACATGGGGACTACAGTC	AATGGAGATGTGTTT	CAGGAGGCTTAACGGGSCCTGCCGA 184
DB	514	GGAGGAGGGCGAGGACAGGAGG	GGGGCAGGAGGGGCGAGGAGG	GGGGCGAGG 573
QY	185	TGGCTTACGCTGCGATTGCC	CAGGCGACAGGTTGACT	CAGGAGCCTGAGAGCATCCGCAA 244
DB	574	GGGGCAGAGGGGCGAGGACGAG	GGGCGAGGAGGGGCGAGGAGG	GGGCGAGG 633
QY	245	GTGGAGAGAGGAGCAGAGAA	GAAGCTCGAGAGTTTGAT	TGCTGTCTCGAAGGTGACCGA 304
DB	634	GGGGCAGGACGAGGAGGGG	CAGGACGAGGAGGGGCGAGG	AGGGGCGAGGAGG 693
QY	305	ACAGGAGTTGGCGGAGAGGGCC	AAAAAGACTTGGAGAGT	TGGACACCGCCCAAGTGA 364
DB	694	GCAGGAGGGCGAGGACAGGAG	GGGGCGAGGAGGGGCGAGG	AGGGGCGAGGAGG 753
QY	365	ACAGGTTGAGAAGAAACAGAT	CAACAACAGGGGCATCGGA	AGAGGCTTTTGTGAAAGAATC 424
DB	754	GGGGCAGGACGAGGAGGGG	CAGGAGGGGCGAGGAGG	GGGGCGAGGAGG 813
QY	425	CAAGGAGGAGACCC	CAGGCGACAGATGGGAG	AAAGGTG 461
DB	814	GGAGGAGGGCGAGGAGGGG	CAGGAGGAGGAGGAGG	GGGG 850

RESULT TO

US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 611411
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Hiang, Betty
; APPLICANT: Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; TITLE OF INVENTION: System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/050,863
; FILING DATE: 30-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-09-050-863-2

Query Match	9.5%	Score 50.6;	DB 3;	Length 2580;
Best Local Similarity	46.9%;	Pred. No. 0.00075;		
Matches 158;	Conservative 0;	Mismatches 179;	Indels 0;	Gaps 0;
QY	125	GGAGGACATGGGGNCTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCTGCGCA	184	
DB	897	GGAGGAGGGCGAGGACGAGGAGGCGGCGAGGAGGGGCGAGGACGAGGAGGAGGGGCGAGGA	956	
QY	185	TGGCTACGCTGCGATTCCACGCGCGACAGTGTGACTTCAGGAGCGCTCAGAGCATCCGCAA	244	
DB	957	GGGGCAGAGGGGCGAGGACGAGGAGGAGGGGCGAGGACGAGGAGGGGCGAGGA	1016	
QY	245	GTGAGAGAGGAGCAGAGAAGAAAAAGCTGCGAGGAGTTGGATCTGCTCGAAGTGACCGA	304	
DB	1017	GGGSCAGGAGCAGGAGGGGCGAGGACGAGGAGGAGGGGCGAGGAGGGGCGAGGA	1076	
QY	305	ACAGGAGTGGCGGAGAGGCCAAAAAGACCTTGGAGGAGTGGAAACGAGCGCCAAAGTGA	364	
DB	1077	GCAGGAGGGGCGAGGACGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGA	1136	
QY	365	ACAGTGTGAGAAGAAACAAGATCAACAACAGGGCATCGGAAAGAGGCTTTTGTGAAGAATC	424	
DB	1137	GGGGCAGGACAGGAGGGGCGAGGAGGGGCGAGGACGAGGAGGGGCGAGGAGGAGCA	1196	
QY	425	CAAGGAGAGACCCGAGGCACAGAGTGGGAGAGGTG	461	
DB	1197	GGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGGG	1233	

RESULT 11

US-09-359-081-2
; Sequence 2, Application US/09359081
; Patent No. 6316223
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; Payan, Don
; Hiang, Betty
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/359,081
; FILING DATE: 22-Jul-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/050,863
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-359-081-2
Query Match 9.5%; Score 50.6; DB 3; Length 2580;
Best Local Similarity 46.9%; Pred. No. 0.00075;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
Qy 125 GGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCTTCCCGA 184
Db 897 GGAGGGGGCAGGAGCAGGAGGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 956
Qy 185 TGGTACGCTGCGATTGCCCGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCCGCA 244
Db 957 GGGGCGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1016
Qy 245 GTGAGAGAGGAGCAGAGAGAGAGAGGCTCAGGAGTTGGATGCTGCTCGAAGGTGACCGA 304
Db 1017 GGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1076
Qy 305 ACAGGAGTGGGGGAGAGAGGCGCAAAAGAGACCTGGAGGAGTGGAGCCAGCGCCAAAGTGA 364
Db 1077 GCAGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1136
Qy 365 ACAGTTGAGAGAGACAGATCAACACAGGGCATCGAAGAGGCTTTTGTGAAGAATC 424
Db 1137 GGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1196
Qy 425 CAAGGAGGAGACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1258

Db 1197 GGAGGAGGCGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1233

RESULT 12

US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: Prom Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEENA
US-09-130-114-1
Query Match 9.5%; Score 50.6; DB 2; Length 5452;
Best Local Similarity 46.9%; Pred. No. 0.00097;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
Qy 125 GGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCTTCCCGA 184
Db 1908 GGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1849
Qy 185 TGGTACGCTGCGATTGCCCGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCCGCA 244
Db 1848 GGGGCGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1789
Qy 245 GTGAGAGAGGAGCAGAGAGAGAGAGGCTCAGGAGTTGGATGCTGCTCGAAGGTGACCGA 304
Db 1788 GGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1729
Qy 305 ACAGGAGTGGCGGCGAGAGGCGCAAAAGAGACCTGGAGAGTGGAAACCAAGCCCAAGTGA 364
Db 1728 GCAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669
Qy 365 ACAGTTGAGAGAGAACAGATCAACACAGGGCATCGAAGAGGCTTTTGTGAAGAATC 424
Db 1668 GGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1609
Qy 425 CAAGGAGGAGACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 461
Db 1608 GGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1572
RESULT 13
US-09-647-344A-14/c
; Sequence 14, Application US/09647344A
; Patent No. 6586180
; GENERAL INFORMATION:
; APPLICANT: Ruffner, Duane E.
; APPLICANT: Pierce, Michael L.
; APPLICANT: Chen, Zhidong
; TITLE OF INVENTION: Directed Antisense Libraries
; FILE REFERENCE: T6678.PCT.US
; CURRENT APPLICATION NUMBER: US/09/647,344A
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: PCT/US99/06742
; PRIOR FILING DATE: 1999-03-28
; NUMBER OF SEQ ID NOS: 50
; SEQ ID NO 14
; LENGTH: 8705
; TYPE: DNA
; ORGANISM: Artificial Sequence

US-09-620-925-1

Query Match		9.5%;	Score 50.6;	DB 3;	Length 9600;
Best Local Similarity		46.9%;	Pred. No. 0.0012;		
Matches 158;		Conservative 0;	Mismatches 179;	Indels 0;	Gaps 0;
Qy	125	GGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTGCCGA	184		
Db	943	GGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGA	1002		
Qy	185	TGGCTACGCTGCGATTGCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCCGCAA	244		
Db	1003	GGGCGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA	1062		
Qy	245	CTGGAGAGAGGAGCAGAAAGGCTGCAGGAGTTGGATGCTGCTCGAAGGTGACCGA	304		
Db	1063	GGGCGAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGA	1122		
Qy	305	ACAGGAGTGGCGGAGAGGCCAATAAGACCTGGAGGAGTGGAAACAGCCGCCAAAGTGA	364		
Db	1123	GCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGA	1182		
Qy	365	ACAGGTTGAGAGAACAGATCAACAACAGGSCATCGGAAGAGGCTTTTGTGAAGATC	424		
Db	1183	GGGCGAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGAGGAGGGGCAGGAGGGGCAGGAGCA	1242		
Qy	425	CAAGGAGGAGACCCAGGACACAGAGTGGGAGAGGTG	461		
Db	1243	GGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG	1279		

Search completed: January 5, 2006, 21:35:23
Job time : 162 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 5, 2006, 19:33:09 ; Search time 804 Seconds

(without alignments)
5502.633 Million cell updates/sec

Title: US-10-664-705-145

Perfect score: 535

Sequence: 1 cggcgctcttcggccag.....cgcctgcctgggtgctcat 535

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	535	100.0	535	US-10-664-705-145	Sequence 145, App
2	471	88.0	982	US-09-917-800A-1407	Sequence 1407, App
3	471	88.0	982	US-10-388-934-227	Sequence 227, App
4	471	88.0	982	US-10-191-803-265	Sequence 265, App
5	471	88.0	982	US-10-152-319A-2036	Sequence 2036, App
6	451.8	84.4	1051	US-10-439-703-14	Sequence 14, Appl
7	451.8	84.4	1051	US-10-240-425-1304	Sequence 1304, App
8	410.8	76.8	636	US-10-956-157-3009	Sequence 3009, App
9	410.8	76.8	636	US-10-956-157-8244	Sequence 8244, App
10	387.8	72.5	1134	US-10-240-425-1303	Sequence 1303, App
11	380.8	71.2	894	US-09-823-245A-178	Sequence 178, App
12	379.2	70.9	907	US-09-823-245A-178	Sequence 178, App
13	364.4	68.1	490	US-09-918-995-20102	Sequence 20102, A
14	293.8	54.9	493	US-09-918-995-26855	Sequence 26855, A
15	268.8	50.2	485	US-09-918-995-27337	Sequence 27337, A
16	213.4	39.9	1164	US-10-264-049-328	Sequence 328, App
17	174.2	32.6	400	US-10-242-535A-10186	Sequence 10186, A
18	174.2	32.6	400	US-10-085-783A-10186	Sequence 10186, A
19	138.6	25.9	1105	US-09-880-107-2289	Sequence 2289, App
20	137.2	25.6	459	US-09-918-995-28203	Sequence 28203, A
21	125.8	23.5	338	US-10-242-535A-11137	Sequence 11137, A
22	125.8	23.5	338	US-10-085-783A-11137	Sequence 11137, A
23	122.2	22.8	419	US-10-242-535A-29408	Sequence 29408, A

24	122.2	22.8	419	7	US-10-085-783A-29408	Sequence 29408, A
25	115.4	21.6	473	3	US-09-918-995-10148	Sequence 10148, A
26	108	20.2	543	4	US-09-925-065A-555818	Sequence 555818, A
27	104.6	19.5	489	7	US-10-424-599-116616	Sequence 116616, A
28	90.2	16.9	612	4	US-09-925-065A-899680	Sequence 899680, A
29	79.2	14.8	476	7	US-10-242-535A-24924	Sequence 24924, A
30	79.2	14.8	476	7	US-10-085-783A-24924	Sequence 24924, A
31	78.4	14.7	522	4	US-09-925-065A-642865	Sequence 642865, A
32	78.2	14.6	1084	10	US-11-097-143-41371	Sequence 41371, A
33	78.2	14.6	3201	10	US-11-097-143-41371	Sequence 41371, A
34	72.6	13.6	484	3	US-09-918-995-20959	Sequence 20959, A
35	72	13.5	1764	9	US-10-450-783-25625	Sequence 25625, A
36	71	13.3	461	7	US-09-918-995-28043	Sequence 28043, A
37	66.6	12.4	411	7	US-10-242-535A-24460	Sequence 24460, A
38	66.6	12.4	411	7	US-10-085-783A-24460	Sequence 24460, A
39	65.2	12.2	658	3	US-09-871-161-327	Sequence 327, App
40	59.2	11.1	6227	6	US-10-240-485-51	Sequence 51, Appl
41	59	11.0	600	9	US-10-972-079-49308	Sequence 49308, A
42	57.2	10.7	439	3	US-09-918-995-10020	Sequence 10020, A
43	57	10.7	628	6	US-10-029-386-22859	Sequence 22859, A
44	57	10.7	3920	10	US-11-097-143-21013	Sequence 21013, A
45	55.8	10.4	659158	3	US-09-771-208-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1

US-10-664-705-145
; Sequence 145, Application US/10664705
; Publication No. US20040152107A1
; GENERAL INFORMATION:
; APPLICANT: Altar, Anthony C.
; APPLICANT: Laeng, Pascal
; APPLICANT: Young, Theresa A.
; APPLICANT: Charles, Vinod
; APPLICANT: Bukhman, Yury
; APPLICANT: Jurata, Linda
; TITLE OF INVENTION: GENE SIGNATURE OF ELECTROSHOCK THERAPY AND METHODS OF USE
; FILE REFERENCE: 03235/100M087-US2
; CURRENT APPLICATION NUMBER: US/10/664,705
; CURRENT FILING DATE: 2003-09-18
; PRIOR APPLICATION NUMBER: US 60/411,718
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/431,882
; PRIOR FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 60/479,970
; PRIOR FILING DATE: 2003-06-18
; NUMBER OF SEQ ID NOS: 152
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 145
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-664-705-145

Query Match 100.0%; Score 535; DB 7; Length 535;

Best Local Similarity 100.0%; Pred No. 1.2e-144; Indels 0; Gaps 0;
Matches 535; Conservative 0; Mismatches 0;

QY	1	CGCGCGCCTTCTCGGCCAGCAGGAGCGAGATTGTCGCATCGAGATGACTCGGGTT	60
DB	1	CGCGCGCCTTCTCGGCCAGCAGGAGCGAGATTGTCGCATCGAGATGACTCGGGTT	60
QY	61	TCGGGCGACCTCGCCGACCGCAGTCTCGCCAGCCCGGACTCGCGAGCGGGGTTG	120
DB	61	TCGGGCGACCTCGCCGACCGCAGTCTCGCCAGCCCGGACTCGCGAGCGGGGTTG	120
QY	121	GTTCGGAGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACCGGCGCTG	180
DB	121	GTTCGGAGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACCGGCGCTG	180
QY	181	CCGATGGCTACGCTCGGATTGCCCGGACAGGTTGACTCAGGAGCCTTGAGAGCATCC	240


```
; PRIOR APPLICATION NUMBER: US 60/330,462
; PRIOR FILING DATE: 2001-10-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2036
; LENGTH: 982
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. NM_053835
US-10-152-319A-2036

Query Match      88.0%; Score 471; DB 7; Length 982;
Best Local Similarity 90.8%; Pred. No. 4.7e-126;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 CGGCCGCTTCTCGCCAGCAGGAGCGGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
Db 1 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 99 CGGCCGCTTCTCGCCAGCAGGAGCGGAGATTGCTGGCATCGAGAATGACTCGGGTT 158
Db 99 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 TCGGGGACCTTCCGCGCAGCAGGAGCGGAGATTGCTGGCATCGAGAATGACTCGGGTT 120
Db 61 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 159 TCGGGGACCTTCCGCGCAGCAGGAGCGGAGATTGCTGGCATCGAGAATGACTCGGGTT 218
Db 159 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GTTCGGAGGACATGGGAGTACAGTCAATGAGATGTGTTTCAGAGGCTTAACCGGGCTG 180
Db 121 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 219 GTTCGGAGGACATGGGAGTACAGTCAATGAGATGTGTTTCAGAGGCTTAACCGGGCTG 278
Db 219 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 CCGATGGCTAGCTGCGATTGCCAGCGGACAGGTTGACTCAGAGGCTTACCGAGCATCC 240
Db 181 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 279 CCGATGGCTAGCTGCGATTGCCAGCGGACAGGTTGACTCAGAGGCTTACCGAGCATCC 338
Db 279 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GCAAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 241 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 339 GCAAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 398
Db 339 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 CCGAACAGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 301 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 399 CCGAACAGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
Db 399 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 GTGACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 394
Db 361 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 459 GTGACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 518
Db 459 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 395 -----GGCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
Db 395 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 519 AGCCAGATGCTGATACCATTTGGCTATGTGGCATCGGAGAGAGAGAGAGAGAGAGAGAG 578
Db 519 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 427 AGGAGGAGACCCAGGACAGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486
Db 427 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 579 AGGAGGAGACCCAGGACAGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
Db 579 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 487 AGAGCAGCAACATGTTAAAGACGTTCCCGCTCGGCTCGGCTCAT 535
Db 487 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 639 AGAGCAGCAACATGTTAAAGACGTTCCCGCTCGGCTCGGCTCAT 687
Db 639 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-10-439-703-14
; Sequence 14, Application US/10439703
; Publication No. US20040018527A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Jenny
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Differential Patterns of Gene Expression that Predict for Docetax
; FILE REFERENCE: Chemoresensitivity and Chemoresistance
; CURRENT APPLICATION NUMBER: US/10/439,703
; PRIOR FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: US 60/381,141
; PRIOR FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 91
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1051
; TYPE: DNA
; ORGANISM: Human
US-10-439-703-14

Query Match      84.4%; Score 451.8; DB 7; Length 1051;
Best Local Similarity 90.3%; Pred. No. 1.7e-120;
Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1 CGGCCGCTTCTCGCCAGCAGGAGCGGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
Db 210 CGGCCGCTTCTCGCCAGCAGGAGCGGAGATTGCTGGCATCGAGAATGACTCGGGTT 269
QY 61 TCGGGGACCTTCCGCGCAGCAGGAGCGGAGATTGCTGGCATCGAGAATGACTCGGGTT 120
Db 270 TCGGGGACCTTCCGCGCAGCAGGAGCGGAGATTGCTGGCATCGAGAATGACTCGGGTT 329
QY 121 GTTCGGAGGACATGGGAGTACAGTCAATGAGATGTGTTTCAGAGGCTTAACCGGGCTG 180
Db 330 GTTCGGAGGACATGGGAGTACAGTCAATGAGATGTGTTTCAGAGGCTTAACCGGGCTG 389
QY 181 CCGATGGCTAGCTGCGATTGCCAGCGGACAGGTTGACTCAGAGGCTTACCGAGCATCC 240
Db 390 CTGATGGCTAGCTGCGATTGCCAGCGGACAGGTTGACTCAGAGGCTTACCGAGCATCC 449
QY 241 GCAAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 450 GCAAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 509
QY 301 CCGAACAGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 510 CCGAACAGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 569
QY 361 GTGACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Db 570 GTGACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 629
QY 421 AATCAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db 630 AATCAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 689
QY 481 ACCCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
Db 690 ACCCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 744

RESULT 7
US-10-240-425-1304
; Sequence 1304, Application US/10240425
; Publication No. US2004003502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph E.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1304
; LENGTH: 1051
; TYPE: DNA
```

```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. U5204040033502A1 M20470
US-10-240-425-1304

Query Match      84.4%; Score 451.8; DB 7; Length 1051;
Best Local Similarity 90.3%; Pred. No. 1.7e-120;
Matches 483; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy      1 CGGCGCGCTTCCTGGCCCGCAGCAGGAGCGAGATTCCTGGCATCGAGATGACTCGGGTT 60
Db      210 CGGCGCGCTTCCTGGCCCGCAGCAGGAGCGAGATTCGAGCATAGAGACGACGAGGCT 269

Qy      61 TCGGGGCGACCTGCGCGCGCAGCAGGTGGCTCTCGCGACGCCCGACTCGCGAGCGGGGTG 120
Db      270 TCGGGGCGACCTGCGCGCGCAGCATCGGCCCCCGCGAGCGGGCCCCACGAGTGGGGCTG 329

Qy      121 GTTCGGAGGACATGGGGACTACAGTCAATCGAGATGTGTTTCAGGAGGCTTAAACGGGCTG 180
Db      330 GTTCTGAGGACATGGGGACCAAGTCAATCGAGATGTGTTTCAGGAGGCCACGCTCTTG 389

Qy      181 CCGATGCTTACGCTGCGATTTGCCAGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 240
Db      390 CTGATGCTTACGCGAGCCATTTGCCAGGCTGACAGGCTGACCCAGGAGCCTGAGAGCATCC 449

Qy      241 GCAGTGGAGAGGAGCAGAGAAAGGCTGCAGGAGTTGGATGCTGCTCGAAGGTGA 300
Db      450 GCAAGTGGCAGAGGAGCAGAGGAAACGGCTGCAGAGACTGGATGCTGCTATCAAGGTCA 509

Qy      301 CCGAACAGGAGTGGCGGGAGAGGCCAAAAAGACCTGGAGGAGTGGAAACGAGCGCCAA 360
Db      510 CGGNAACAGGAATGGCGGGAGNAGGCCAAGAAGNACCTGGAGGAGTGGAAACGAGCGCCAGA 569

Qy      361 GTGAACAGGTTGAGAAGACAGATCAACACAGGGCATCGGAAGAGGCTTTTGTGAAG 420
Db      570 GTGAACAGTAGAGAAGAACAGAATCAACAAACCGGGCATCCGAGGAGGCTTTTGTGAAG 629

Qy      421 AATCCAAGGAGGAGACCCGAGGACAGAGTGGGAGAAGGTGGCCCGACTGTGTGACTTCA 480
Db      630 AATCCAAGGAGGAGACCCGAGGACAGAGTGGGAGAAGGTGGCCCGACTGTGTGACTTCA 689

Qy      481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTCGCGCTCGGTGCTCAT 535
Db      690 ACCCAAGAGCAGCAGCAGGTGCAAGAGATGTGTCCCGCTCGCGCTCGGTGCTCAT 744

```

RESULT 8
 US-10-956-157-3009
 ; Sequence 3009, Application US/10956157
 ; Publication No. US20050118625A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wyeth
 ; APPLICANT: Mounts, William
 ; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
 ; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
 ; FILE REFERENCE: 031896-043000 (AM 101081)
 ; CURRENT APPLICATION NUMBER: US/10/956.157
 ; CURRENT FILING DATE: 2004-10-04
 ; NUMBER OF SEQ ID NOS: 319805
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 3009
 ; LENGTH: 636
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-956-157-3009

Query Match 76.8%; Score 410.8; DB 9; Length 636;
 Best Local Similarity 89.5%; Pred. No. 1.2e-108;
 Matches 453; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 31 AGATTGCTGGCATCGAATGACTCGGGTTTCGGGGCACCTGCGCGCCAGCGCATGCGCT 90
 1 AGATTGCGAGGATAGAGAAACACAGAGGCTTCGGGGCACTTCGCGGCAGCGCATGCGGCC 60

	Qy	91	CTGGCGACCCGGA	CTCGAGCGGGGTGGTTC- GGAGGCATCGGGACTACAGTCAAT	144
	Db	61	CCGCGCAGCCGGGCC	CCACGAGTGGGGCTGGTTCTTGAGGACATGGGACCACAGTCAAT	120
	Qy	150	GGAGATGTGTTTT	CAGGAGGGCTAAACGGGCGCTCCCGATGGCTACGTGCGATTGCCCAGGCG	209
	Db	121	GGAGATGTGTTTT	CAGGAGGCCAACGGTCTGCTGATGGCTACGCAGCCATTGCCCAGGCT	180
	Qy	210	GACAGTTGACT	CAGGAGCCTGAGAGCATCCGCAAGTGGAGACGAGGACGAGNAGRAAGG	269
	Db	181	GACAGGCTGAC	CCCCAGGACCTTGAGAGCATCCGCAAGTGGCGAGAGGACGACAGGAAAACGG	240
	Qy	270	CTGCAGAGATTGGAT	TGTCCTCGCAAGGTGACCGAACAAGGAGTGGCGGAGAAAGGCCAAA	329
	Db	241	CTGCAGAGCTGGAT	GTGTCATCTTAAGGTCAAGGNAACAGGAATGGCGGAGAAGGCCAAG	300
	Qy	330	AAAGACCTGGAGGAGT	GGAAACCAAGCGCCAAAGTAGTGAACAGGTTGAGAGAAGAACAGATCAAC	389
	Db	301	AAGGACCTGGAGGAGT	GGAAACCAAGCGCCAGAGTGAAACAAGTAGAGAGAAGAACAGATCAAC	360
	Qy	390	AACAGGGCATCGGA	AGAGGCTTTTGTGAAGAAATCCAAGGAGGAGACCCCAGGCACAGAG	449
	Db	361	AACCGGGCATCCG	AGGAGGCTTTTGGTGAAGGAATCCAAGGAGGAGACCCCAGGCACAGAG	420
	Qy	450	TGGGAGAAGGTGGCC	ACAGCTGTGTGACTTTCAAACCCCTAAGAGCAGCACAGCAATGTATAAGAC	509
	Db	421	TGGGAGAAGGTGGC	CCAGCTATGTGACTTTCAAACCCCAAGAGCAGCACAGCAGTGTCAAAGAT	480
	Qy	510	GTGTCCCGCCCTCG	CTCGGTGCTCAT	535
	Db	481	GTGTCCCGCCCTCG	CTCGGTGCTCAT	506

RESULT 9
US-10-956-157-8244/c
; Sequence 8244, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:

```

? APPLICANT: Wyeth
? APPLICANT: Mounts, William
? TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR
? TITLE OF INVENTION: HUMAN OSTEOARTHRITIS
? TITLE OF INVENTION: HUMAN OSTEOARTHRITIS
? FILE REFERENCE: 031896-043000 (AM 101081)
? CURRENT APPLICATION NUMBER: US/10/956,157
? CURRENT FILING DATE: 2004-10-04
? NUMBER OF SEQ ID NOS: 319805
? SOFTWARE: PatentIn version 3.2
? SEQ ID NO 8244
? LENGTH: 636

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-8244

```

	Query Match	76.8%	Score 410.8	DB 9	Length 636
	Best Local Similarity	89.5%	Prod. No. 1.2e-108		
	Matches 453	Conservative	0	Mismatches 52	Indels 1
				Gaps	1
QY	31	AGATTGCTGGCATCTGAGAAATGACTCGGGTTCGCGGGCACCTGCGCGCCAGCCAGGTGGCGCT	90		
DB	636	AGATTGCAGGCATAGAGAACACGAGGGCTTCGCGGGCACCTGCGCGAGCCATCGGGCCC	577		
QY	91	CTGGCGACCCGGACTCCGAGACGGGGGTGGTTC-GGAGGACATGGGGACTACAGTCAAT	149		
DB	576	CGCGCAGCCGGGGCCCCACGAGTGGGGCTGGTTCTTGAGGACATGGGGACCACAGTCAAT	517		
QY	150	GGAGATGTGTTTCAGGAGGCTTAAAGGGCTCCCGATGGCTACGCTGCGATTGCCCCAGCGC	209		
DB	516	GGAGATGTGTTTCAGGAGGCCAACGGTCTCTGATGGCTACGCAGCCATTTGCCAGGCT	457		
QY	210	GACAGGTTGACTCAGGAGCCTTGAGGACATCCGCAGTGGAGAGGACGACGAGNAAGG	269		
DB	456	GACAGGCTGACCAGGAGCCTTGAGAGCATCCGCAAGTGGCAGAGGACGACGGNAACGG	397		


```
Db 654 GCGAGAGGACGAGGAAACGGCTGCAAGAGCTGGATGCTCATTAAGTCAACGGAACA 595
Qy 308 GGAATGCGGAGAGGCAAAAGAGCTGGAGAGTGGAAACAGCGCCAAAGTGAACA 367
Db 594 GGAATGCGGAGAGGCAAAAGAGCTGGAGAGTGGAAACAGCGCCAAAGTGAACA 535
Qy 368 GGTGGAAGAACAGATCAACACAG----- 394
Db 534 AGTAGAGAGAACAGATCAACACCGGATCGCTGACAAAGCATTTCTACAGAGCCGAGA 475
Qy 395 -----GGCATCGGAAGAGGCTTTTGTGAAGAAATCAAGGAGGA 433
Db 474 TGCTGATATCATCGGCTACGTGGCATCCGAGGAGGCTTTCTGTAAGGAATCCAAGGAGGA 415
Qy 434 GACCCAGGACACAGAGTGGAGAGAGTGGCCAGCTGTGTGACTTCAACCCCTAAGAGCAG 493
Db 414 GACCCAGGACACAGAGTGGAGAGAGTGGCCAGCTGTGTGACTTCAACCCCTAAGAGCAG 355
Qy 494 CAAGCAATGTAAAGACGTGTCCCGCTCGCTCGGTGCTCAT 535
Db 354 CAAGCAGTGCAAAGATGTGTCCCGCTCGCTCGGTGCTCAT 313
```

```
RESULT 12
US-09-823-245A-178/c
; Sequence 178, Application US/09823245A
; Publication No. US20020039760A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fectel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalak
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6401
; CURRENT APPLICATION NUMBER: US/09/823,245A
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/194,941
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 178
; LENGTH: 907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-823-245A-178
```

```
Query Match 70.9%; Score 379.2; DB 3; Length 907;
Best Local Similarity 81.6%; Pred. No. 1.7e-99;
Matches 475; Conservative 0; Mismatches 53; Indels 54; Gaps 1;

Qy 8 CTTCTGCGCCAGCAGGAGGAGAGATTCGTGTCAGATGAGATGACTCGGGTTTCGGGC 67
Db 907 CTTCTGCGCCAGCAGGAGGAGAGATTCGTGTCAGATGAGATGACTCGGGTTTCGGGC 848
Qy 68 ACCTGCGCCAGCAGGAGGAGGAGATTCGTGTCAGATGAGATGACTCGGGTTTCGGGC 127
Db 847 ATCTGCGCCAGCAGGAGGAGGAGATTCGTGTCAGATGAGATGACTCGGGTTTCGGGC 788
Qy 128 GGCATGCGGAGCTACAGTCAATGAGATGTTTTCAGGAGGCTTAAACGGGCTCGCGATGG 187
Db 787 GGCATGCGGAGCTACAGTCAATGAGATGTTTTCAGGAGGCTTAAACGGGCTCGCGATGG 728
Qy 188 CTACGCTGCGATTCGCCAGCGGACAGGTTGACTCAGGAGCTCAGAGATCCGCAAGTG 247
Db 727 CTACGCGAGCATTTGCCAGGCTGACAGGCTGACCCAGGAGCTCAGAGATCCGCAAGTG 668
Qy 248 GAGAGAGGACGAGAAAGAGGCTGCAGGAGTTGGATGCTGCTCGAAGGTGACCGAACA 307
```

```
Db 667 GCGAGAGGACGAGGAAACGGCTGCAAGAGCTGGATGCTCATTAAGTCAACGGAACA 608
Qy 308 GGAATGCGGAGAGGCAAAAGAGCTGGAGAGTGGAAACAGCGCCAAAGTGAACA 367
Db 607 GGAATGCGGAGAGGCAAAAGAGCTGGAGAGTGGAAACAGCGCCAAAGTGAACA 548
Qy 368 GGTGGAAGAACAGATCAACACAG----- 394
Db 547 AGTAGAGAGAACAGATCAACACCGGATCGCTGACAAAGCATTTCTACAGAGCCGAGA 488
Qy 395 -----GGCATCGGAAGAGGCTTTTGTGAAGAAATCAAGGAGGA 433
Db 487 TGCTGATATCATCGGCTACGTGGCATCCGAGGAGGCTTTCTGTAAGGAATCCAAGGAGGA 428
Qy 434 GACCCAGGACACAGAGTGGAGAGAGTGGCCAGCTGTGTGACTTCAACCCCTAAGAGCAG 493
Db 427 GACCCAGGACACAGAGTGGAGAGAGTGGCCAGCTGTGTGACTTCAACCCCTAAGAGCAG 368
Qy 494 CAAGCAATGTAAAGACGTGTCCCGCTCGCTCGGTGCTCAT 535
Db 367 CAAGCAGTGCAAAGATGTGTCCCGCTCGCTCGGTGCTCAT 326
```

```
RESULT 13
US-09-918-995-20102
; Sequence 20102, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20102
; LENGTH: 490
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(490)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-20102
```

```
Query Match 68.1%; Score 364.4; DB 3; Length 490;
Best Local Similarity 89.3%; Pred. No. 3e-95;
Matches 392; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 97 AGCCCGGACTCGCAGCGGGGTGTTTCGAGAGGACATCGGGACTACAGTCAATGGAGATG 156
Db 44 AGCCCGGCCCCACGAGTGGGGTGTCTCTGAGGACATGGGGACACAGTCAATGGAGATG 103
Qy 157 TGTTTCAGGAGGCTAACGGGCTCGGATGGCTGCTACGCTCCGATTCGCCAGGCGGACAGT 216
Db 104 TGTTTCAGGAGGCTAACGGGCTCGGATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 163
Qy 217 TGACTCAGGAGCTCAGAGATCCCGCAAGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 276
Db 164 TGACCCAGGAGCTCAGAGATCCCGCATGTGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 223
Qy 277 AGTTGGATGCTGCTCGAAGGTGACCGGACAGGATGCGCGGAGAGGAGGAGGAGGAGGAG 336
Db 224 AGCTGATGCTGCTCATCTTAAGGTTCACGGAAACAAAGATGGCGGAGAGGAGGAGGAGGAG 283
Qy 337 TGGAGAGTGGAAACAGCGCCAAAGTGAACAGGTTTGAAGAGAACAGATCAACACAGGAG 396
Db 284 TGGAGAGTGGAAACAGCGCCAGAGAGTGAACAAAGTGAAGAAACAGATCAACACAGGAG 343
```



```
Qy 397 CATCGAAGAGGCTTTTGTGAAGAAATCCAAAGGAGGAGACCCAGGACACAGTGGGAGA 456
|||
Db 344 CATCGAGGAGGCTTTTGTGAAGAAATCCAAAGGAGGAGACCCAGGACACAGTGGGAGA 403
|||
Qy 457 AGGTGGCCACGCTGTGTGACTTCAACCCCTAAAGAGCAGCAAGCAATGTAAAGACGTGTCCC 516
|||
Db 404 AGGTGGCCACGCTGTGTGACTTCAACCCCTAAAGAGCAGCANGCAGTGCATAGATGTGTCCC 463
|||
Qy 517 GCCTGCGCTCGGTGCTCAT 535
|||
Db 464 GCCTGCGCTCGGTGCTCAT 482
|||

RESULT 14
US-09-918-995-26855
; Sequence 26855, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26855
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(493)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-26855
```

```
Query Match 54.9%; Score 293.8; DB 3; Length 493;
Best Local Similarity 80.7%; Pred. No. 8.3e-75;
Matches 343; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 39 GGCATCGAGAATGACTCGGGTTTCGGGGCACTCGCGCAGCAGGTGGCTCTGCGCAG 98
|||
Db 61 GGCATAGAGAACGACGAGGGCTTCGGGGCACTCGCGCAGCAGGTGGCTGGCGCAG 120
|||
Qy 99 CCCGACTCGCGAGCGGGGTGGTTTCGGAGGACATGGGGAATACAGTCAATGGAGATGTG 158
|||
Db 121 CCGGGCCCCACGAGTGGGGCTGTCTGAGGACATGGGACCAACAGTCAATGGAGATGTG 180
|||
Qy 159 TTTGAGGAGCTAACGGGCTTCGCGATGGCTACGCTGCGATTTGCCAGGCGGACAGGTTG 218
|||
Db 181 TTTGAGGAGGCAACGGTCTCTGCTGATGGCTACGCGAGCATTTGCCAGGCTGACAGGCTG 240
|||
Qy 219 ACTCAGGAGCTGAGAGCATCCGAAGTGGAGAGGAGGAGCAAGAAAGGCTGCAGGAG 278
|||
Db 241 ACCAGGAGCTGAGAGCATTCGACAGTGGCGAGAGGAGGAAACGCTGCAAGAG 300
|||
Qy 279 TTGGATGCTGCTCGAAGGTACCGAACAGAGGTGGCGGAGAAAGGCCAAAAAGACTTG 338
|||
Db 301 CTGGATGCTGCATCTTAAGGTCAACGAAACAGGAATGGCGGGAGAAAGGCCAAGAGACCTG 360
|||
Qy 339 GAGGAGTGGACACGCGCCAAAGTGAACAGGTTGAGAGAACAGATCAACACAGGGCA 398
|||
Db 361 GAGGAGTGGAAACGAGCGCCAGAGTGAACAGTAGAGAGAAACAAGATCAACACCGGATC 420
|||
Qy 399 TCGGAAGAGGCTTTTGTGAAGAAATCCAAAGGAGGAGACCCAGGACACAGTGGGAGAG 458
|||
Db 421 GCTGACAAAGCAATTCACAGAGCCAGATGCTGATATCATCGGCTACGTGGCATCCGAG 480
|||
Qy 459 GTGGC 463
|||
Db 481 GAGGC 485
```

RESULT 15

```
US-09-918-995-27337
; Sequence 27337, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27337
; LENGTH: 485
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(485)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27337
```

```
Query Match 50.2%; Score 268.8; DB 3; Length 485;
Best Local Similarity 77.1%; Pred. No. 1.4e-67;
Matches 327; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 40 GCATCGAGAATGACTCGGGTTTCGGGGCACTCGCGCAGCAGGTGGCTCTGCGCAGC 99
|||
Db 58 GCATACAGAACGACAAAGGCTTCTGGGACCTCGCCGTGAGCCATGCAGTTCCTCCGCGCAT 117
|||
Qy 100 CCGGACTCGCGAGCGGGGTGGTTTCGGAGGACATGGGGACTTACAGTCAATCGAGATGTGT 159
|||
Db 118 CGGGCCCCACGAGTGGGGCTGGGGGTGAGGACATGGGGACCAACAGTCAATGGAGATGTGT 177
|||
Qy 160 TTGAGGAGGCTAACGGGCTTCGCGATGGCTACGCTACGCTGCGATTTGCCAGGCGGACAGTTGA 219
|||
Db 178 TTGAGGAGGCGCAACGGTCTCTGCTGATGGCTACGCGAGCCATTTGCCCAGGCTGACAGGCTGA 237
|||
Qy 220 CTGAGGAGCCTGAGAGCATCCGAAGTGGAGAGGAGCAGAAAGGCTGCAGGAGT 279
|||
Db 238 CCCATGAGCCTGAGAGCATTCGCAAGTGGCGAGAGGAGCAGAGAAACGCTGCAAGAGC 297
|||
Qy 280 TGGATGCTGCTCGCAAGGTGACCCGAACAGGAGTGGCGGAGAAAGGCCAAAAAGACCTGG 339
|||
Db 298 TGGATGCTGCATCTAAGGTCAACGAAACAGGAATGGCGGAGAAAGGCCAAGATGACCTGG 357
|||
Qy 340 AGGAGTGGAAACGAGCGCCAAAGTGAACAGGTTGAGAGAAACAAGATCAACACAGGGCAT 399
|||
Db 358 AGGAGTGGAAACGAGCTGCTAGTGAACAAGTTTGAAGAAACAAGATCAACAAACCGGATCG 417
|||
Qy 400 CGGAAGAGGCTTTTGTGAAGAAATCCCAAGGAGGAGACCCAGGACACAGTGGGAGAAAGG 459
|||
Db 418 CTGACAAAGCAATTCACAGCAGCCAGATGCTGATATCATCGGCTACGTGGCATCCGAGG 477
|||
Qy 460 TGGC 463
|||
Db 478 AGGC 481
```

Search completed: January 5, 2006, 21:49:02
Job time : 807 secs


```

; GENERAL INFORMATION:
; APPLICANT: PUSZTAI, LAJOS
; APPLICANT: SYMANS, W. FRASER
; APPLICANT: HESS, KENNETH R.
; APPLICANT: AYERS, MARK
; APPLICANT: STEC, JAMES
; TITLE OF INVENTION: MULTIGENE PREDICTORS OF RESPONSE TO CHEMOTHERAPY
; FILE REFERENCE: UTXC:890US
; CURRENT APPLICATION NUMBER: US/10/955,054A
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 94
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-955-054A-94

Query Match      39.9%; Score 213.4; DB 6; Length 1023;
Best Local Similarity 63.7%; Pred. No. 4e-47;
Matches 341; Conservative 0; Mismatches 191; Indels 3; Gaps 1;

QY 1 CGCGCGCTTCTCGCCAGCAGGAGAGAGAGATGCTGGCATCGAATGACTCGGGTT 60
DB 177 CTGCGGCTTCTTGGCGCAGCAAGAGAGAGAGATGCGGGCATCGAGAAACGACGAGGCT 236

QY 61 TCGGGGACCTCGCCGACAGCAGGTGGCTCTGCGCAGCCCGGACTCGGAGCGGGGTG 120
DB 237 TCGCCATCTTGGACGCGGCGC---GCCCGCGGCGCCGCGCAGCGCGCGCGGGG 293

QY 121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGGCTG 180
DB 294 GTCCGGATGCTGTTGATGGATTAATGAATGGTGAATATCTACAGGAAGTAAATGTCCAA 353

QY 181 CGGATGGCTTACCTCGATTTGCCAGCGGAGCAGAGTTGACTCAGGAGCCTGAGAGCATCC 240
DB 354 CAGACAGTTATCGAGCTATTTTCAAGTGTGATCGATTGCGAGCAGGCTGAAAGTATCC 413

QY 241 GCAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 414 GTAAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 473

QY 301 CGGACAGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 474 AAGAAGCAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 533

QY 361 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 534 ACGAGCAGCTACAGAAACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 593

QY 421 AATCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 594 ACATTGACGAGTGTCTCCAGGAGCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 653

QY 481 ACCCTAAGCAGCAGCAGCAGTGAAGAGAGTGTCCCGCTCGGCTCGGCTCAT 535
DB 654 ACCCCAAAGTCTAGCAAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 708

RESULT 5
US-11-136-527-347
; Sequence 347, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4443
; LENGTH: 1169
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-4443

Query Match      27.3%; Score 146; DB 7; Length 1169;
Best Local Similarity 56.8%; Pred. No. 2.5e-29;
Matches 259; Conservative 14; Mismatches 180; Indels 3; Gaps 1;

QY 1 CGCGCGCTTCTTCTGGCCAGCAGGAGAGAGAGATGCTGGCATCGAATGACTCGGGTT 60
DB 210 CCGCGGCTTCTTCTGGCCAGCAGGAGAGAGATGCTGGCATCGAATGACTCGGGTT 269

QY 61 TCGGGGACCTCGCCGACGAGCAGGTGGCTCTGCGCAGCCCGGACTCGGAGCGGGGTG 120
DB 270 TCGCCATCTTAGACGGGCGC---GCCCGCGGCGCGAGMACACGGCGAGCGCGGGG 326

QY 121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTTAACGGGCTG 180
DB 327 GTCTGTATGCGGTTGATGGAGTGAATGGGAATACTACAGGAGAGCAATGTGTCCAA 386

QY 181 CCGATGGCTACGTCGCGATTTGCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 240
DB 387 CAGACAGTTACGCGAGCCATTTTCAAGGTGGATCGGTTGCGAGTCAAGAGCTGAAAGTATCC 446

QY 241 GCAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 447 GTAAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 506

QY 301 CGGACAGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 507 AGGAAGCAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 566

QY 361 GTGAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 567 ATGACAGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 626

QY 421 AATCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 456
DB 627 RASCTKGMWAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662

RESULT 6
US-11-136-527-4443
; Sequence 4443, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4443
; LENGTH: 1169
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-4443

Query Match      27.3%; Score 146; DB 7; Length 1169;
Best Local Similarity 56.8%; Pred. No. 2.5e-29;
Matches 259; Conservative 14; Mismatches 180; Indels 3; Gaps 1;

QY 1 CGCGCGCTTCTTCTGGCCAGCAGGAGAGAGAGATGCTGGCATCGAATGACTCGGGTT 60
DB 210 CCGCGGCTTCTTCTGGCCAGCAGGAGAGAGATGCTGGCATCGAATGACTCGGGTT 269

```


